THE BASES OF PROBABILITY[1]

B. O. KOOPMAN

The subject for consideration today forms an aspect of a somewhat venerable branch of mathematical theory; but in essence it is part of a far older department of thought—the ancient science of logic. For it is concerned with a category of propositions of a nature marked by features neither physical nor mathematical, but by their rôle under the aspect of the reason. Their essential characteristic is their involvement of that species of relation between the knower and the known evoked by such terms as probability, likelihood, degree of certainty, as used in the parlance of intuitive thought. It is our threefold task to transcribe this concept into symbols, to formulate its principles, and to study its properties in their inner order and outward application.

As prelude to this undertaking it is necessary to set forth certain conventions of logic. Propositions are the elements of symbolic logic, but they may play the rôle of *contemplated propositions* (statements in quotation marks) or of *asserted propositions* (statements regarded as true throughout a given manipulation or deduction); and it is necessary to take account of this in the notation for the logical constants. We shall employ the symbols for negation ($\sim$), conjunction or logical product ($\cdot$), and disjunction or logical sum ($\vee$), and regard them as having no assertive power: they combine contemplated propositions into contemplated propositions and asserted propositions into asserted propositions *of the same logical type.* Quite other shall be our convention regarding implication ($\subset$) and equivalence ($=$): they combine contemplated propositions into asserted propositions, and shall not be used to combine asserted propositions in our present study. If a and b stand for contemplated propositions, the assertion that a is false (true) shall be written $a=0$ ($a=1$), and the assertion that a implies b, $a \subset b$ or $a \sim b = 0$; it is thus quite different from the contemplated proposition $\sim(a \sim b)$. Finally it is universally asserted that $a \sim a = 0$, $a \vee \sim a = 1$, and in fact all the laws of Boolean algebra are regarded as assertions. We shall assume their elements to be fa-

[1] An address delivered before the New York meeting of the Society on February 24, 1940, by invitation of the Program Committee.

For the details of the theory here expounded, see the two publications of the present author. *The axioms and algebra of intuitive probability*, Annals of Mathematics, (2), vol. 41 (1940), pp. 269–292 (herein to be abbreviated as AAP) and *Intuitive probability and sequences* (forthcoming in the Annals of Mathematics) (abbreviation PS).

miliar, and shall accept without question the intuitive logical background implied in their manipulation and interpretation.[2]

The technical logician will observe that there is here involved a meta-mathematical question. But space forbids us to introduce and explain here the modern terminology of this subject, and compels us to throw a perhaps undue burden on the terms "contemplated" and "asserted proposition."

The next step towards our goal consists in defining a category of propositions which are to form the substratum upon which the structure of our theory is to be erected. We will designate by *experimental propositions* such statements of the outcome of a particular physical or biological event as may in principle be verified by the performance of a single crucial experiment. Thus "it will rain on this roof tomorrow at this hour" or "Mr. X made a mistake in his accounts last Monday" are experimental propositions, whereas "Newtonian mechanics is correct" is not: the motion of bodies can always be accounted for by assuming sufficiently complicated laws of force in Newton's equations, so its truth, while experimental in meaning, is not determined by a crucial experiment but rather by its ability to harmonize the results of many such experiments in an acceptably simple manner. Experimental propositions shall be denoted by lower case Latin letters; and inasmuch as a finite set of crucial experiments may always be regarded as constituting a single crucial experiment, finite combinations of letters by means of $(\sim \cdot \vee)$ also denote experimental propositions. Finally, experimental propositions shall be regarded as contemplated propositions.[3]

At this point the category of propositions may be introduced which forms the subject of the theory of probability. We will take as the conceptual germ from which the whole theory springs the ordering of two events in the relation "not more probable than," a relation to be denoted by the partial ordering symbol $(\prec)$—which like $(\subset, =)$ shall have assertive force. One could of course develop a theory of such assertions as $a \prec b$; but it would prove insufficient for the purposes of probability, as one needs for example to compare the probability of a assuming h true with its probability assuming h false. The definitive form of propositions sought is the following, in which a, b, h, k are experimental propositions and $h \neq 0$, $k \neq 0$:

[2] For references see E. V. Huntington, Transactions of this Society, vol. 35 (1933), pp. 274–304.

[3] In its occurrence in a/h the experimental proposition h is in a sense "temporarily asserted," i.e., a is viewed on the assumption that h is true. But we are applying the term asserted proposition only to those held as true on *both* sides of $a/h \prec b/k$.

a on the presumption h is no more probable than b on the presumption k

and this is symbolized as $a/h \prec b/k$ (or equivalently, $b/k \succ a/h$). Thus from the four contemplated experimental propositions and the logical symbol ($/ \prec /$) an asserted proposition $a/h \prec b/k$ (to be called a ($\prec$) proposition) is formed which is to constitute the building stone of the whole theory of probability.

Before proceeding further, two difficulties must be surmounted.

Firstly it may be objected that $a/h \prec b/k$ is a proposition of such vague and subjective order that it may not merely be held by one person and rejected by another, but that one and the same individual may sometimes assent to it and at other times and in a different mood reject it. If this is so, how in the nature of things can such propositions form the subject matter of a precise mathematical science? Secondly it may be objected that the probability of a proposition a depends on a body of knowledge going far beyond the fact that h is true: It will involve propositions of higher logical types such as the laws of logic—and of probability itself; and perhaps even matters of subconscious moods, associations, artistic taste, and the like.

We believe that these two objections are answered at one stroke by adhering to the following convention, or rather, clarification of the use and laws of ($\prec$) propositions. *A given individual at a given moment* may be regarded as assenting to a certain set of ($\prec$) propositions; ignoring what he may hold at any other moment or what others may believe, that set of ($\prec$) propositions which he holds at that given moment must have certain relations with one another which may be called *relations of consistency*. It is to their formulation and study that we conceive the present science to be devoted. So viewed, the analogue with strict logic is clear: many may disagree with me when I assert $a = 0$; but every one whose mind is constructed on normal lines will agree that if $a = 0$ then $(\sim\sim a) = 0$. Along with the first objection, this convention answers the second, for by positing a given individual at a given moment in the consideration of any set of ($\prec$) propositions, the body of knowledge becomes fixed throughout, and so does not require explicit symbolization.

A third objection which might be voiced is against the restriction of the application of probability to experimental propositions. Why can we not compare the probability of two physical theories, for example? In answer to this we can say only that with the present restriction many grave logical difficulties are avoided and a theory is obtained which covers all the classical cases of mathematical proba-

bility and many others as well; and our limitation of the scope is a matter of practical strategy rather than of principle. It would undoubtedly be of interest to extend the present ideas to propositions of higher logical types, classes of propositions, logical systems and the like.

There is evidently no difference in intuitive meaning between a/h and ah/h (or, dually, $a \vee \sim h/h$), and we will make the notational convention that any one of these symbols be replaceable by any other. This has an interesting algebraic counterpart. Let $\mathfrak{A}$ be the Boolean ring determined by all the experimental propositions considered in a given discussion.[4] To make the presumption that $h \in \mathfrak{A}$ is true (equivalently, that $\sim h$ is false) is to make the presumption that any two propositions a and b of $\mathfrak{A}$ for which $a \sim b \subset \sim h$ and $b \sim a \subset \sim h$ are equivalent, i.e., both true or false simultaneously. But this "identification" of all so-related pairs a, b is precisely the formation of the quotient ring $\mathfrak{A}/(\sim h)$ whose elements $a/(\sim h)$ are the remainder classes with respect to the principle ideal $(\sim h)$. It is purely for convenience that we write a/h in lieu of $a/(\sim h)$. Thus if $\mathcal{A}$ is the class of all remainder classes in $\mathfrak{A}$ with respect to all its principal ideals, the $(\prec)$ symbol introduces a partial ordering of the elements of $\mathcal{A}$.[5]

We are now ready to undertake our second task and lay down the axioms which govern any aggregate of $(\prec)$ propositions. It will be noted that they all have $(\subset, =)$ or $(\prec)$ propositions as hypothesis and as conclusion, and that in each case where the conclusion is a non-trivial $(\prec)$ proposition, this is true of the hypothesis as well. Finally, a tacit assumption is always made: *no denominator* $= 0$.

The axioms

V. Axiom of verification. *If* $k \subset b$, *then* $a/h \prec b/k$.

I. Axiom of implication. *If* $a/h \prec b/k$ *and* $h \subset a$, *then* $k \subset b$.

R. Axiom of reflexivity. *If* $h = k$ *and* $ah = bk$, *then* $a/h \prec b/k$.

T. Axiom of transitivity. *If* $a/h \prec b/k$ *and* $b/k \prec c/l$, *then* $a/h \prec c/l$.

A. Axiom of antisymmetry. *If* $a/h \prec b/k$, *then* $\sim a/h \succ \sim b/k$.

C. Axioms of composition. *Let* $0 \neq a_1 \subset b_1 \subset c_1$ *and* $0 \neq a_2 \subset b_2 \subset c_2$.

[4] For Boolean rings and their ideals, see M. H. Stone, Transactions of this Society, vol. 40 (1936), pp. 37–111.

[5] One should guard against the notion that this ordering has any simple relation with the ordering of the remainder classes with respect to class inclusion.

C_1. *If* $a_1/b_1 \prec a_2/b_2$ *and* $b_1/c_1 \prec b_2/c_2$, *then* $a_1/c_1 \prec a_2/c_2$.
C_2. *If* $a_1/b_1 \prec b_2/c_2$ *and* $b_1/c_1 \prec a_2/b_2$, *then* $a_1/c_1 \prec a_2/c_2$.

D. AXIOMS OF DECOMPOSITION (QUASI-CONVERSES OF C). *Let* $0 \neq a_1 \subset b_1 \subset c_1$, $0 \neq a_2 \subset b_2 \subset c_2$, *and* $a_1/c_1 \prec a_2/c_2$. *Then if either symbol in* (i): $(a_1/b_1, b_1/c_1)$ *has the* $(\succ)$ *relation with either in* (ii): $(a_2/b_2, b_2/c_2)$, *then the remaining symbol in* (i) *has the* $(\prec)$ *relation with that in* (ii). (*Thus* D *contains four axioms.*)

P. AXIOM OF ALTERNATIVE PRESUMPTION. *Let* $a/hb \prec r/s$ *and* $a/h \sim b \prec r/s$; *then* $a/h \prec r/s$.

S. AXIOM OF SUBDIVISION. *For each positive integer n the following axiom is posited*:
S_n. *If* $a_1 \vee \cdots \vee a_n = a \neq 0$, $b_1 \vee \cdots \vee b_n = b \neq 0$, $a_i a_j = b_i b_j = 0$ *for all* $i \neq j$, *and lastly if*

$$a_1/a \prec a_2/a \prec \cdots \prec a_n/a,$$

$$b_1/b \succ b_2/b \succ \cdots \succ b_n/b,$$

then $a_1/a \prec b_1/b$.

The Axioms V, R, T, A are simply the transcription into the present language of facts so familiar as scarcely to require comment. The partial ordering property of $(\prec)$ expresses itself by R and T; it leads to the definition of equiprobability $a/h \approx b/k$, inferior probability $a/h \prec b/k$ and incomparability $a/h \| b/k$ in the usual manner. The question of whether one can go further and assume that the entities a/h form the elements of a lattice with respect to $(\prec)$ will naturally be raised; until now we have been unable to make any use of this idea, and if our experience is borne out we shall be in the presence of the first non-trivial example of a partially ordered set which is not a lattice.

Axiom I is in sharp contrast with the familiar circumstance that the numerical probability of an event may be unity (i.e., the same as a certain event) without that event's being certain. This is because numerical probability gives but a blurred rendering of the ultimate logical relations between probability and certainty.

As for Axiom C and its converse D, the following verbal rendering of C_1 may be given: If a_1 depends for its possibility on b_1, and likewise a_2 on b_2, and if c_1 is less likely to lead to b_1 than is c_2 to b_2, and if finally b_1 is in turn less likely to lead to a_1 than is b_2 to a_2, then c_1 is less likely to lead to a_1 than c_2 to a_2. So stated, it exhibits a sort of inner transitivity. All the other cases have a corresponding phraseology.

One might be disposed to regard Axiom P as a theorem which could be proved by arguing that since the hypothesis $a/hb \prec r/s$, $a/h{\sim}b \prec r/s$ tells us that a on the presumption h is not more likely than r/s both when b is true and when b is false, it must be so in all cases, i.e., $a/h \prec r/s$. Carrying this species of reasoning a little further, we could prove that such a relation as $a/h \approx {\sim}a/h$ is impossible: Since the assertion $(a \vee {\sim}a) = 1$ is always made we would conclude that the only possibilities are $a = 1$ or ${\sim}a = 1$ (i.e., $a = 0$); in either case $a/h \approx b/k$ is impossible—hence it is never possible. In essence this is the old objection of the elementary student who voices it by saying that since an event will either happen or not happen, it is absurd to say that its probability of happening could ever be $\frac{1}{2}$. We all know how to answer him by general reference to the dependence of probability on a body of knowledge;[6] but we are in a position here to give the answer in a precise logical form: The fallacy lies in confusing the assertion $(a \vee {\sim}a) = 1$ with the assertion "$a = 1$ *or* ${\sim}a = 1$" [which might be written $(a = 1) \vee ({\sim}a = 1)$]. *The distinction between an asserted disjunction and a disjoined assertion is fundamental*: $(u \vee v) = 1$ *must never be confused with* $(u = 1) \vee (v = 1)$. The disregard of this distinction has led to more difficulties in the foundations of probability than is often imagined. It is now clear that the above proof of Axiom P is fallacious since it confuses $(b \vee {\sim}b) = 1$ with $(b = 1) \vee ({\sim}b = 1)$. As a matter of fact the same proof would have provided an infinite extension of Axiom P, an extension which leads to paradoxes.[7]

Axiom S is epitomized in the idea that if a first event is less likely of occurrence than its opposite and if a second is more likely than its opposite, then the first is less likely than the second. While we have not succeeded in simplifying the general case (beyond restricting S_n to prime values of n), we still feel confident that those more skilful than ourselves may have better success.

As a purely formal matter it may be remarked that in a system *completely* ordered by $(\prec)$, Axioms P and S are logical consequences of the rest.

Our second task being complete, we pass to the third, the deduction of all the useful theorems of probability from the axioms. But before proceeding it may be remarked that we have traversed the path of all mathematical disciplines: One proposes to study a subject of which

[6] The notion that the uncertainty resides in the events themselves rather than in the mind of the individual contemplating them, the appeal to the "principle of uncertainty," etc., betrays merely a misconception both of probability and of quantum mechanics.

[7] Cf. PS, §2.

one is made aware through the intuition, the senses, and such non-mathematical modes of perception. In undertaking it one introduces symbols and statements of axioms and laws in terms of these. But from this point on everything (except the interpretation) becomes purely mathematical: The symbolic abstractions become the ultimate objects of study, the axioms and laws become the postulates. Thenceforth we may discard the intuitionalistic introduction of our symbols and axioms and regard the latter as pure conventions (postulates) pertaining to the former, which are taken as "undefinables."

The further developments fall into three groups of theorems: the theorems on comparison, the theorems on numerical probability, and the theorems on statistical weight or frequency in a sequence.

In the first group we shall confine ourselves to citing the following typical ones. No comments appear necessary.

THEOREM. *If* $ah \neq 0$ *or* h, *then* $0/1 < a/h < 1/1$.

THEOREM. *If* $a_1/h_1 \prec a_2/h_2$, $b_1/h_1 \prec b_2/h_2$, *and* $a_1b_1h_1 = a_2b_2h_2 = 0$, *then* $a_1 \vee b_1/h_1 \prec a_2 \vee b_2/h_2$.

THEOREM. *If* $a/hc_i \prec r/s$ $(i = 1, \cdots, n)$ *and* $hc_ic_j = 0$ *for all* $i \neq j$, *then* $a/hc \prec r/s$ *where* $c = c_1 \vee \cdots \vee c_n$.

The second group starts with the introduction of the numerical probability $p(a/h) = p(a, h)$, i.e., the number between 0 and 1 forming the basis of the classical theory. This is accomplished as follows:

DEFINITION. *Any set of propositions* $(u_1, \cdots, u_n)$ *shall be called an* n*-scale when they satisfy the conditions* (i) $u_1 \vee \cdots \vee u_n = u \neq 0$; (ii) $u_iu_j = 0$ (*all* $i \neq j$); (iii) $u_i/u \approx u_j/u$ (*all* i, j).

ASSUMPTION. *Any positive integer* n *being given, the conceptional existence of at least one* n*-scale may be assumed.*

This is the only principle which need be assumed in addition to the axioms in all the further developments of the theory.[8] It is of a fundamentally different nature from the axioms and might be compared with the assumption so familiar in thermodynamics and other parts of physics of the possibility of a conceptual experiment.

THEOREM. *If* $(u_1, \cdots, u_n)$ *is an* n*-scale and* $(v_1, \cdots, v_m)$ *an* m*-scale,*

$$u_1 \vee \cdots \vee u_\nu/u \; <, \approx, > \; v_1 \vee \cdots \vee v_\mu/v$$

according as $\nu/n \; <, =, > \; \mu/m$.

[8] In the exhibition of certain paradoxes the extension of the assumption regarding the existence of n-scales to $n = \aleph_0$ is required.

Let $t(n)$ be the maximum value of t for which $u_1 \vee \cdots \vee u_t/u \prec a/b$ holds ($t=0$ corresponding to $0/u \prec a/b$) and $T(n)$ the minimum value of T for which $a/b \prec u_1 \vee \cdots \vee u_T/u$. By the previous theorem $t(n)$ and $T(n)$ are independent of the particular n-scale chosen; for fixed a/b they are always defined functions of n. We then prove the following:

THEOREM. *The limits* $p_*(a/h) = \lim_{n\to\infty} t(n)/n$ *and* $p^*(a/h) = \lim_{n\to\infty} T(n)/n$ *always exist, and* $0 \leqq p_*(a/h) \leqq p^*(a/h) \leqq 1$.

The numbers $p_*(a/h)$ and $p^*(a/h)$ may be called the lower and upper numerical probabilities of a/h.

DEFINITION. *If* $p_*(a/h) = p^*(a/h)$, a/h *is said to be appraisable and to have* $p(a/h) = p_*(a/h) = p^*(a/h)$ *as its numerical probability.*

All the classical theorems follow. We give merely the following two examples:

THEOREM. *Let* $a \subset b \subset c$; *if* a/c *and* b/c *are appraisable and if* $p(b/c) \neq 0$, *then* a/b *will be appraisable and* $p(a/c) = p(a/b)p(b/c)$.

THEOREM. *Let* a/h *and* b/h *be appraisable. Then* $a \vee b/h$ *will be appraisable if and only if* ab/h *is appraisable and it will then follow that*

$$p(a/h) + p(b/h) = p(a \vee b/h) + p(ab/h).$$

If now we consider the Boolean ring determined by the totality of propositions considered in a given discussion and assume that every a/h formed in it is appraisable, we are at the threshold of the classical theory. For we are in possession of an additive function obeying all the postulates required for its derivation. Its rôle is thus revealed as a theory of (unfaithful) numerical representation of relations belonging to the more far-reaching logical theory.

Before passing to the third group of results the question as to complete additivity is in order: Our axioms establish only the restricted additivity of numerical probability. The example of the infinite sequence of propositions $a_1, a_2, \cdots$ for which $a_1 \vee a_2 \vee \cdots = 1$, $a_i a_j = 0$ $(i \neq j)$ and $p(a_1/1) = p(a_2/1) = \cdots$ shows that the equation

$$1 = p(a_1 \vee a_2 \vee \cdots /1) = p(a_1/1) + p(a_2/1) + \cdots$$

is impossible.[8] This could be interpreted either by regarding the assumptions concerning $a_1, a_2, \cdots$ as self-contradictory or regarding them to be valid and holding the view that complete additivity is not a general property, but occurs only in an important class of special cases where its validity is a consequence of the physical circum-

stances and not of the logical aspect of probability. We have adopted the latter position.

Anyone conversant with modern theoretical physics is aware of the fundamental rôle played therein by probability. Statistical mechanics is a familiar example; but even more significant is the case of quantum mechanics, the laws of which can not be stated except in terms of probability. Now if these sciences are to be regarded as affording objective pictures of nature, how can their laws involve in an essential manner the notion of probability if this is indeed a concept of logic—a mode of thought? The answer to this question is immediate: the "probability" of these branches of physics is a misnomer for statistical weight or frequency in a sequence. If an event E in such a theory can have two possible outcomes, "success" (labeled 1) and "failure" (labeled 0), a conceptually infinite sequence of trials under "the same conditions" furnishes an infinite sequence of zeros and ones (α): $(\alpha_1, \alpha_2, \cdots)$ $(\alpha_n = 0, 1)$. The physical assumption that $w = \lim_{n\to\infty} (\alpha_1 + \cdots + \alpha_n)/n$ exists is made and this statistical weight or frequency w is what is designated by the word "probability" of success of E. But the whole objective content of the physical laws in question involves solely the notion of frequency.

Yet the intuitive conception of probability upon which the present work is based plays an essential part in connection with frequency. Its rôle becomes manifest at that very moment when the experimental significance of w is sought—significance, that is, to a pre-named individual in terms of the only phenomena which can come within his ken. Then it is that we become aware that a link is needed between the finite sets of trials—all that we can actually observe—and the mathematical idealization of frequency.[9] Analysis reveals that the only possible link is bound to involve the intuitive idea of probability.[10] Granting then the present theory, are we enabled to solve the problem? That we are indeed able to give a complete and precise solution and to do so without assuming any further principles is the content of the third group of theorems, to which we now turn.

---

[9] This remains true even when frequency is thought of as a ratio in a finite sequence containing a larger number of trials than can come before the individual's observation. This is the difficulty which confronts any attempt to dispense with everything of the essence of intuitive probability and replace it by a theory of frequency. It is an attempt often made with the object of freeing the science of subjectivism (sic) and of retaining therein only an account of that which scientists "really observe"; by a strange irony it places the theory of probability completely out of contact with what any given human being could ever observe.

[10] Cf. PS, §1.

Let a_n be the experimental proposition "E succeeds at the nth trial" (so that $\alpha_n=1$). Let h denote the statement of the common experimental condition at the instance of each trial. Then the following theorem is typical.

THEOREM. *Hypothesis* 1: $\lim_{n\to\infty}(\alpha_1+\cdots+\alpha_n)/n=w$. *Hypothesis* 2: *For each positive integer* t

$$a_{i_1}\cdots a_{i_t}/h \approx a_{j_1}\cdots a_{j_t}/h,$$

where $(i_1,\cdots,i_t)$ *is any set of* t *distinct positive integers and* $(j_1,\cdots,j_t)$ *a similar set. Conclusion*: a_i/h *is appraisable and* $p(a_i/h)=w$.

The first idea which should enter the mind of the mathematician is that the sequence (α) can (at least when $0<w<1$) be reordered so as to yield a different frequency; yet apparently this must still be equal to w; is this not contradictory? The answer consists in examining the precise logical meaning of Hypothesis 1. Firstly, let $W(w,\mu,n)$ denote the assertion: *h implies that the number of true propositions in the set* $(a_1,\cdots,a_n)$ *is between* $n(w-1/\mu)$ *and* $n(w+1/\mu)$. Then Hypothesis 1 becomes: For any given integer μ there exists an m such that for all $n\geqq m$ assertion $W(w,\mu,n)$ is made. In logical symbols this is the familiar

$$\prod_{\mu=1}^{\infty}\sum_{m=1}^{\infty}\prod_{n=m}^{\infty}W(w,\mu,n).$$

This is all clear enough; the ambiguity appears when the logical form for assertion $W(w,\mu,n)$ is sought, for it turns out that there are many. The following is in a certain sense the weakest; it is the one for which the above theorem is proved; it is suitable for relating frequency to probability in physics:

$$W(w,\mu,n): \qquad h\subset\sum_{(p,q)}a_{p_1}\cdots a_{p_t}\sim a_{q_1}\cdots\sim a_{q_f}.$$

Here the $\sum$ calls for the disjunction of all terms where $(p_1,\cdots,p_t,q_1,\cdots,q_f)$ are all possible sets of $t+f$ distinct integers between 1 and n and where t is the least integer $\geqq n(w-1/\mu)$ and f the least integer $\geqq n(1-w-1/\mu)$. But Hypothesis 1 with this form of $W(w,\mu,n)$ is not the one which makes it possible to reorder the sequence so as to produce the contradiction. For this purpose it is necessary to replace it by the entirely different $W'(w,\mu,n)$:

$$W'(w,\mu,n): \qquad \sum_{(p,q)}(h\subset a_{p_1}\cdots a_{p_t}=a_{q_1}\cdots=a_{q_f}).$$

Thus the paradox is resolved as in earlier cases by maintaining the distinction between an assertion of a disjunction and a disjunction of assertions.

The complementary rôle of the hypotheses of this theorem is worthy of note. Hypothesis 1 exhausts the purely objective state of affairs, while Hypothesis 2 which expresses a sort of intuitive random quality, contains the assumption of a subjective nature which allows an actual living being to capture the otherwise inaccessible objective fact and relate it to his own world of possible experience by the agency of intuitive probability. Thus the theorem renders unto objective reality that that is objective, and unto the subjective intuition that which pertains thereto.

No discussion of the bases of probability would be complete at the present day which did not make reference to alleged cases of the violation of certain principles of classical probability by the phenomena of quantum mechanics. The precise form at which we have here arrived makes it particularly simple to subject every such case to minute scrutiny. While we have no time for examples here, we are publishing elsewhere a discussion which shows that it is the physical circumstances to which the laws of probability apply and never the laws themselves which are altered.[11] It would indeed be hard to imagine how it could be otherwise. For insofar as the laws of probability are laws of thought, they are prior to experimental verification in the laboratory. For how indeed can such experiments prove any statement? Firstly, when the statement is an experimental proposition, then a crucial experiment suffices: but this is evidently not the case for the axioms of probability, which are not experimental propositions. Secondly, when the statement introduces harmony and intelligibility into an ensemble of statements proved in the laboratory: but the axioms of probability appear rather in the rôle of the *criteria* of such harmony and intelligibility. To argue, finally, that the axioms repose on *subjective* experiments and hence are experimental in character is beside the point since we are considering quantum mechanics which is based on experiments on electron tubes and things of this sort which are hardly in a class with the subjective experiments by means of which we become aware of our own rational processes.

Having dwelt so long on the positive side, it behooves us to mention a fundamental limitation of these results. The theory cannot prove that the probability of heads on the toss of a coin is $\frac{1}{2}$. More generally, it is as impotent to derive a non-trivial ($<$) proposition from a

[11] Cf. PS, §6.

set of propositions of whatever character not containing a ($\prec$) proposition (stated or implied) as are the laws of Newtonian mechanics to predict the position of a particle at a given time when no initial conditions are assigned. For after all, the whole theoretical structure is but the statement *in extenso* of the laws of consistency governing an aggregate of ($\prec$) propositions.

The question is naturally raised whether some further principle of a purely formal-logical nature can be enunciated which will establish ($\prec$) propositions *ab ovo*. Having searched high and low in the literature we become aware that every apparent case of such a principle either contains in some veiled form a ($\prec$) proposition in its hypothesis, or else leads to insurmountable paradoxes, as in the case of the principle of sufficient reason or symmetry of ignorance which has so long sullied the name of a priori intuitive probability. The quest for the first ($\prec$) proposition is epitomized by the attempt to devise an experiment proving the irrelevance of some external condition A in a trial of an event E. Such a statement of irrelevance is of course a ($\prec$) proposition. In order to reason that A is irrelevant to E on one occasion from the results of experiments performed on another, one must assume that certain other unavoidable differences between the two occasions are themselves irrelevant to the situation. The difficulty, exactly contrary to Napoleon's Guard, always retreats but does not expire.

It is in the light of experience such as this that we may well ask whether it is not a principle of epistemology itself that blocks our path; and, even as those who having sought in vain for perpetual motion ended by making a virtue of their failure, so we may hazard the view that in principle the authority for the first ($\prec$) proposition does not reside in any general law of probability, logic, or experimental science. And the notion presents itself that such primary and irreducible assumptions are grounded on a basis as much of the aesthetic as of the logical order.

COLUMBIA UNIVERSITY

ALMOST CYCLIC ELEMENTS AND SIMPLE LINKS OF A CONTINUOUS CURVE[1]

F. B. JONES

Many of the definitions and results concerning connected im kleinen continua become useful only when these continua are locally compact. This is especially true in the cyclic element theory. For if a continuous curve M is not locally compact, it is not necessarily true that a simple closed curve in M belongs to a cyclic element of M. Furthermore, if a continuous curve M is not locally compact, it is not necessarily true that a simple closed curve in M belongs to a simple link in M. However, if M is a continuous curve, there are in M certain subcontinua which strongly resemble both cyclic elements and simple links such that if J is a simple closed curve in M, one of them contains J. It is the purpose of this paper to define these sets, to develop a few of their properties, and to show how they are of considerable interest in spaces where the Jordan curve theorem holds true.

1. **Results for complete Moore spaces.** In this section it is assumed that S, the set of all points, is a complete Moore space, that is, Axioms 0 and 1 of R. L. Moore's *Foundations of Point Set Theory*[2] hold true in S.

DEFINITION. *Suppose that K is a nondegenerate subset of a continuous curve M such that* (1) *if A and B are distinct points of K, there exists a simple closed curve lying in M and containing $A+B$, and* (2) *if X is a point of $M-K$, there is some point O of K such that no simple closed curve lying in M contains both X and O. The set K is said to be a "cyclic nucleus"[3] of M.*

---

[1] Presented to the Society, December 28, 1939.

[2] American Mathematical Society Colloquium Publications, vol. 13, New York, 1932. Hereinafter, this book will be referred to as *Foundations*. The reader is referred to *Foundations* for the definition of terms used, but not specifically defined, here.

[3] A continuous curve is defined to be a connected im kleinen continuum. It need *not* be locally compact. It is easy to see with the help of Theorems 118 and 120 in Chapter I and the arguments for Theorems 6 and 7 in Chapter II of *Foundations* that if a nondegenerate continuous curve M is regarded as a space and the term "region" is interpreted to mean a connected open subset of M, then with respect to this interpretation of "point" and "region," Axioms 0, 1, and 2 of *Foundations* hold true in M and "limit point" is invariant under this change. Hence, it is possible to apply certain theorems found in Chapter II of *Foundations* and elsewhere to continuous curves. For example, any two points of a connected open subset D of a continuous curve are the extremities of an arc lying wholly in D.

THEOREM 1. *If K is a cyclic nucleus of a continuous curve M, and AB is an arc which lies in M and whose end points belong to K, then AB is a subset*[4] *of K.*

PROOF. Suppose that O is any point of K and that X is a point of AB. There exist in M simple closed curves J_A and J_B containing $O+A$ and $O+B$ respectively. Let H denote $AB+J_A+J_B$. It is easy to see that H is a compact continuous curve and that no point separates O from X in H. By Theorem 40 on page 124 of *Foundations*, H contains a simple closed curve containing $O+X$. Hence X belongs to K.

THEOREM 2. *If K is a cyclic nucleus of a continuous curve M, and AB is an arc which lies in M and whose end points belong to $\overline{K}$, then $AB-(A+B)$ is a subset of K.*

PROOF. Suppose that X is a point of $AB-(A+B)$. There exist two mutually exclusive connected open subsets D_A and D_B of M which contain A and B respectively such that neither contains X. Since both D_A and D_B contain points of K, D_A+AB+D_B contains an arc $A'B'$ whose end points A' and B' belong to $K \cdot D_A$ and $K \cdot D_B$ respectively. But $A'B'$ must contain X. Hence, by Theorem 1, X belongs to K.

THEOREM 3. *If K is a cyclic nucleus of a continuous curve M, and J is a simple closed curve lying in M and containing more than one point of $\overline{K}$, then J is a subset of K.*

Theorem 3 follows at once from Theorem 2.

THEOREM 4. *If K is a cyclic nucleus of a continuous curve M, then neither K nor $\overline{K}$ contains a cut point of itself.*

Theorem 4 may be established with the help of Theorem 3.

THEOREM 5. *If K_1 and K_2 are cyclic nucleae of a continuous curve M, and $\overline{K}_1$ and $\overline{K}_2$ are identical, then K_1 and K_2 are identical.*

PROOF. If X is a point of K_1, X is a subset of a simple closed curve J lying in K_1. Hence J is a subset of $\overline{K}_2$. By Theorem 3, J is a subset of K_2, and X belongs to K_2. Likewise, any point of K_2 is a point of K_1.

---

[4] A subset K of a continuous curve M which contains every arc in M whose end points belong to K has been called a *basic set* of M. See W. L. Ayres, *Concerning the arc-curves and basic sets of a continuous curve*, Transactions of this Society, vol. 30 (1928), pp. 567–578, and *Concerning the arc-curves and basic sets of a continuous curve, second paper*, Transactions of this Society, vol. 31 (1929), pp. 595–612.

THEOREM 6. *If K is a cyclic nucleus of a continuous curve M, and C is a component of $M-\overline{K}$, then C has at most one boundary point in $\overline{K}$.*[5]

PROOF. It follows from Theorem 2 on page 89 of *Foundations* that C is a connected open subset of M. Suppose that C has two boundary points in $\overline{K}$. Using the theorem just referred to, it is easy to see that M contains an arc AB such that (1) $AB-(A+B)$ is a subset of C and (2) both A and B belong to $\overline{K}$. By Theorem 2, $AB-(A+B)$ belongs to K, which is a contradiction.

THEOREM 7. *If K is a cyclic nucleus of a continuous curve M, and X is a point of $M-K$, then there is at most one point O of K such that some simple closed curve in M contains $O+X$.*

PROOF. If X is not a point of $\overline{K}$, the conclusion of Theorem 7 follows from Theorem 6. If X is a point of $\overline{K}$ and a simple closed curve in M contains X and a point O of K, then by Theorem 3, X belongs to K, which is a contradiction.

THEOREM 8. *If K is a cyclic nucleus of a continuous curve M, and D is a connected open subset of M containing a point of K, then both $D\cdot K$ and $D\cdot\overline{K}$ are connected.*[6]

PROOF. Suppose that either $D\cdot K$ or $D\cdot\overline{K}$ is not connected. Then it is the sum of two mutually separate sets, K_1 and K_2. Let AB denote an arc in D from a point of K_1 to a point of K_2. By Theorem 2, $AB-(A+B)$ is a subset of K. This involves a contradiction.

THEOREM 9. *If K is a cyclic nucleus of a continuous curve M, $\overline{K}$ is a nondegenerate continuous curve which contains no cut point of itself.*

Theorem 9 may be proved with the help of Theorems 4 and 8.

THEOREM 10. *If K_1 and K_2 are cyclic nucleae of a continuous curve M, then either K_1 and K_2 are identical or $\overline{K}_1$ and $\overline{K}_2$ have at most one point in common.*

PROOF. Suppose that $\overline{K}_1$ contains two points of $\overline{K}_2$. Then it is evident from Theorems 4 and 6 that $\overline{K}_1$ is a subset of $\overline{K}_2$. Conversely, $\overline{K}_2$ is likewise a subset of $\overline{K}_1$. Hence $\overline{K}_1$ and $\overline{K}_2$ are identical. By Theorem 5, K_1 and K_2 are identical.

[5] Cf. G. T. Whyburn, *Cyclicly connected continuous curves*, Proceedings of the National Academy of Sciences, vol. 13 (1927), p. 33, Theorem 2.

[6] If the word "open" is omitted from the statement of Theorem 8, the resulting proposition is false. Cf. G. T. Whyburn, *Concerning the structure of a continuous curve*, American Journal of Mathematics, vol. 50 (1928), p. 191, Theorem 30.

THEOREM 11. *If J is a simple closed curve lying in a continuous curve M, then there exists one and only one cyclic nucleus of M containing*[7] *J.*

PROOF. Let G denote the collection of all subsets k of M such that (1) k contains J and (2) every two points of k belong to a simple closed curve in M. Let K denote G^* (that is, the sum of the elements of G) and let A and B denote two distinct points of J. If X and Y are two distinct points of K, there exist in M simple closed curves J_{AX}, J_{BX}, J_{AY}, and J_{BY} which contain $A+X$, $B+X$, $A+Y$, and $B+Y$ respectively. Then $J_{AX}+J_{BX}+J_{AY}+J_{BY}$ is a compact continuous curve no point of which separates X from Y; hence it contains a simple closed curve C containing $X+Y$. Consequently K is an element of G.

Suppose that K is not a cyclic nucleus of M. Then $M-K$ contains a point X such that if O is a point of K, M contains a simple closed curve J_{OX} containing both O and X. Again let A and B denote distinct points of J, and let J_{AX} and J_{BX} denote simple closed curves lying in M and containing $A+X$ and $B+X$ respectively. Then $k=J+J_{AX}+J_{BX}$ is a compact continuous curve containing no cut point of itself, and every two of its points belong to a simple closed curve in k which is a subset of M. Hence k is an element of G. But this involves a contradiction. Therefore K is a cyclic nucleus of M which contains J. By Theorem 10, no other cyclic nucleus of M contains J.

THEOREM 12. *In order that a set K be a cyclic nucleus of a continuous curve M, it is necessary and sufficient that K consist of two points A and B which belong to a simple closed curve in M together with all other points X such that* (1) *some simple closed curve lying in M contains $A+X$ and* (2) *some simple closed curve lying in M contains $B+X$.*

PROOF. It is easily seen from Theorem 7 and the definition of a cyclic nucleus of a continuous curve that the condition is necessary. The condition is also sufficient. For let J denote a simple closed curve in M containing $A+B$ and let H denote the cyclic nucleus of M which contains J. Then H is a subset of K. But it is clear from Theorem 7 that K is a subset of H. Hence K is a cyclic nucleus of M.

THEOREM 13. *If A, B and X are three points of a cyclic nucleus K*

---

[7] Cf. G. T. Whyburn, *Cyclicly connected continuous curves*, loc. cit., p. 34, Theorem 3.

of a continuous curve M, then there exists in K a simple closed curve containing $A+B+X$.

PROOF. Let J denote a simple closed curve in K containing $A+B$. If X does not belong to J, let J_A and J_B denote simple closed curves in M containing $A+X$ and $B+X$ respectively. Suppose that neither J_A nor J_B intersects J in more than one point. Let AB denote an arc of J from A to B. The set $AB+J_A+J_B$ is a compact continuous curve no point of which separates A from B. By Theorem 40 of page 124 of *Foundations*, $AB+J_A+J_B$ contains a simple closed curve C containing $A+B$. Evidently C intersects each of the curves J_A and J_B in more than one point. So in any case there exist a simple closed curve J_1 in M containing $A+B$ and a simple closed curve J_2 in M containing X such that J_2 intersects J_1 in more than one point. Let T denote the component of $J_2-J_1\cdot J_2$ which contains X. Since T is an arc segment, it is easy to see that J_1+T contains a simple closed curve J_3 containing $A+B+X$. By Theorem 3, J_3 is a subset of K.

THEOREM 14. *In order that a set K of more than two points be a cyclic nucleus of a continuous curve M, it is necessary and sufficient that K consist of two points A and B together with all points X such that some simple closed curve lying in M contains $A+B+X$.*

Theorem 14 follows from Theorems 12 and 13.

DEFINITIONS. *A nondegenerate continuous curve M is said to be "almost cyclicly connected" provided that, if A and B are distinct points of M, and R_A and R_B are regions containing A and B respectively, then M contains a simple closed curve containing both a point of R_A and a point of R_B. A subset H of a continuous curve M is said to be an "almost cyclic element" of M provided that* (1) *H is either a cut point or an end point of M or* (2) *H is a nondegenerate almost cyclicly connected continuous curve which is a subset of M but which is not a proper subset of any other almost cyclicly connected continuous curve which is a subset[8] of M.*

THEOREM 15. *If K is a cyclic nucleus of a continuous curve M, then $\overline{K}$ is a nondegenerate almost cyclic element of M.*

PROOF. By Theorem 9, $\overline{K}$ is a nondegenerate continuous curve. Since every two points of K lie together in a simple closed curve which is a subset of K, it is evident that $\overline{K}$ is almost cyclicly con-

[8] Cf. G. T. Whyburn, *Concerning the structure of a continuous curve*, loc. cit., p. 167.

nected. Furthermore, it is clear from Theorem 6, that $\overline{K}$ is not a proper subset of an almost cyclicly connected continuous curve lying in M. Hence $\overline{K}$ is an almost cyclic element of M.

THEOREM 16. *If H is a nondegenerate almost cyclic element of a continuous curve M, then M contains one and only one cyclic nucleus K such that $\overline{K}=H$.*

PROOF. Let J denote a simple closed curve lying in H. By Theorem 11, M contains one and only one cyclic nucleus K containing J. If H is not a subset of $\overline{K}$, then it follows from Theorem 6 that H contains a cut point of itself. This is impossible. Hence H is a subset of $\overline{K}$. But by Theorem 15, $\overline{K}$ is an almost cyclicly connected continuous curve. Consequently, H is not a proper subset of $\overline{K}$.

THEOREM 17. *No two almost cyclic elements of a continuous curve M have more than one point in common.*

THEOREM 18. *If J is a simple closed curve lying in a continuous curve M, then one and only one almost cyclic element of M contains the curve J.*

DEFINITIONS. *Suppose that P is a point of a continuum M and there do not exist two points A and B of M such that* (1) *P separates A from B in M and* (2) *P is the only point of M which separates A from B in M. Let K denote the set of all points X of M such that no point separates P from X in M. Then K will be called a "simple link of M" and P will be called a "proper point of M."*[9]

THEOREM 19. *Every nondegenerate simple link of a continuous curve M is an almost cyclic element*[10] *of M.*

PROOF. Suppose that K is a nondegenerate simple link of M. Let A and B denote two distinct points of K and let AXB denote an arc in M from A to B containing a point X distinct from A and B. The arc AXB belongs to K, and X does not separate A from B in K. Hence there exists in $M-X$ an arc $A'B'$ having only its end points A' and B' in AXB. Let J denote the simple closed curve contained in $AXB+A'+B'$. Obviously J is a subset of K. By Theorem 18, M contains one and only one almost cyclic element H containing J. Since no two points of K are separated in M by any point of K, it follows from Theorems 6 and 16 that K is a subset of H. On the other hand, K contains a point P together with all other points X of M such that

[9] See pages 63 and 72 of *Foundations*.

[10] Cf. Theorem 68 on page 148 of *Foundations*.

X is not separated in M from P by a point of M. Since H contains P and no cut point of itself, it is a subset of K. Consequently K is identical with H.

THEOREM 20. *Every simple link of a continuous curve is itself a continuous curve.*[11]

THEOREM 21. *If a nondegenerate almost cyclic element H of a continuous curve M contains a proper point of M, then H is a simple link*[12] *of M.*

PROOF. If H contains a proper point P of M, then by Theorem 94 on page 68 of *Foundations* M contains one and only one simple link K containing P. Evidently every point of H belongs to K. It follows from Theorems 17 and 19 that H and K are identical.

EXAMPLE. *An almost cyclic element of a continuous curve M need not contain a proper point of M.* Suppose that E is an euclidean 3-space. Let H denote all points $(X, Y, 0)$ of E such that $X^2+Y^2=1$ and let W denote all points (X, Y, Z) of E such that $X^2+Y^2>1$. Now let M denote $H+W$ and define "region" in M as follows: (1) if R is a region in E containing a point of H, then $R\cdot M$ shall be called a "region" in M and (2) if (X, Y, Z) is a point of W and R is a circular region of the plane $Yx-Xy=0$ containing (X, Y, Z) such that $\overline{R}$ contains no point of H, then $R\cdot M$ shall be called a "region" in M. The space M satisfies Axioms 0–2 of *Foundations*. Consequently M is a continuous curve lying in a complete Moore space. It is clear that H is an almost cyclic element of M which contains no proper point of M.

With the theorems and proofs of the preceding pages in mind the reader will find it possible to establish many results for *almost cyclic* elements of a continuous curve in a complete Moore space analogous to those in the literature[13] for *cyclic* elements of a continuous curve in an euclidean n-space by slight changes in the original arguments. For certain results, such as those concerning the number of nondegenerate cyclic elements, it is obviously necessary in complete Moore spaces to assume that the continuous curves involved are separable. One theorem of considerable interest is as follows: *Suppose that M is a continuous curve in a complete Moore space. If the almost cyclic elements of M are regarded as "points" and two such "points" p and q are*

[11] Cf. Theorem 66 on page 147 of *Foundations*.

[12] Cf. Theorem 69 on page 149 of *Foundations*.

[13] Especially those of the references which have been given in the footnotes of this paper.

regarded as contiguous if and only if one of the two continua p and q is a point of the other, then the set of all such "points" is an acyclic continuous curve.[14]

2. **Spaces in which the Jordan curve theorem holds true.** The next two theorems are obtained by a simple application of the preceding results to spaces in which the Jordan curve theorem holds true.

THEOREM 22. *Suppose that S is a space satisfying Axioms* 0, 1, 2, *and* 4 *of Foundations and that H is a nondegenerate almost cyclic element of S. If H is regarded as a space and the term "region" is interpreted to mean a connected open subset of H, then, with respect to this interpretation of "point" and "region," Axioms* 0–4 *of Foundations are satisfied and "limit point" is invariant under this change.*[15]

PROOF. By Theorem 9 on page 96 of *Foundations*, Axioms 0–2 are satisfied in H and "limit point" is invariant. It follows from Theorems 4 and 16 that Axiom 3 is satisfied. If J is a simple closed curve lying in H, then $S-J$ is the sum of two connected domains, E and I, each having J for its boundary. It follows from Theorems 8 and 16 that $H\cdot E$ and $H\cdot I$, if not vacuous, are connected open subsets of H. If some point P of J is not a limit point of $H\cdot E$, then some segment T of J contains P but no point of $H\cdot\overline{E}$. There exists an arc segment W lying in E whose end points lie in T. By Theorems 2 and 16, W is a subset of H. Hence W is in $H\cdot E$, and the end points of W belong to both $H\cdot\overline{E}$ and T, which is a contradiction. Consequently, every point of J is a limit point of $H\cdot E$. It is now evident that J is the boundary with respect to H of $H\cdot E$. Likewise J is the boundary with respect to H of $H\cdot I$. Hence Axiom 4 holds true in H.

THEOREM 23. *Suppose that S is the set of all points and that Axioms* 0, 1, 2, *and* 4 *of Foundations hold true. Then a nondegenerate simple link of S is a nondegenerate almost cyclic element of S and conversely.*

PROOF. It is obvious from Axiom 4 that no point of a simple closed curve is a cut point of S. Hence every point of a simple closed curve

---

[14] Cf. R. L. Moore, *Fundamental theorems concerning points sets*, the Rice Institute Pamphlet, vol. 23 (1936), pp. 1–74; p. 74 in particular.

[15] AXIOM 2. *If P is a point of a region R, there exists a nondegenerate connected domain containing P and lying wholly in R.*

AXIOM 3. *If O is a point, $S-O$ is connected.*

AXIOM 4. *If J is a simple closed curve, $S-J$ is the sum of two mutually separated connected point sets such that J is the boundary of each of them.* (The Jordan curve theorem.)

is a proper point of S. The conclusion of Theorem 23 now follows immediately from Theorems 19 and 21.

I wish to point out that Theorems 22 and 23 shed some light on the role played by Axiom 3 in the sequence of axioms in *Foundations*. A space satisfying Axioms 0–4 of *Foundations* is identical with its one almost cyclic element. Hence, by assuming Axiom 3 in addition to Axioms 0, 1, 2, and 4 one has merely confined one's investigation to a single almost cyclic element of the space. In fact with the preceding theorems in mind it is easy to see that many of the theorems in the literature which hold true in spaces satisfying Axioms 0–4 also hold true in spaces satisfying only Axioms 0, 1, 2, and 4. This is true, for instance, of all the theorems in Chapter III of *Foundations* except Theorems 0, 2, 21, 23, and 24.

THE UNIVERSITY OF TEXAS

THE CONFORMAL NEAR-MOEBIUS TRANSFORMATIONS[1]

EDWARD KASNER AND JOHN DE CICCO

1. Introduction. In a previous paper,[2] we discussed the point transformations of the plane with reference to the maximum number of circles preserved. A nonconformal point transformation of the complex plane converts at most $2\infty^2$ circles into circles.[3] A conformal transformation, not of the Moebius type, carries at most $2\infty^1$ circles into circles (excluding the $2\infty^1$ minimal lines which become minimal lines). A Moebius transformation carries the entire family of ∞^3 circles into circles. From these results, we obtain the following two characterizations of the group of Moebius transformations: (1) if $3\infty^2$ circles are carried into circles after a point transformation, then the same is true for all circles, and the point transformation is therefore a Moebius transformation; and (2) any conformal transformation which converts $3\infty^1$ circles into circles is a Moebius transformation.

In this paper, we shall determine the set of all *conformal near-Moebius transformations*. That is, we shall obtain the set of all conformal transformations which convert *exactly* $2\infty^1$ circles into circles. *Any conformal near-Moebius transformation is of the form* M_2TM_1 *where* M_1 *and* M_2 *are Moebius transformations and* T *is any one of the three transformations* e^z, $\log z$, z^n. The two families preserved are two orthogonal pencils of circles.

The conformal near-collineation problem[4] is a special case of our problem. *Any conformal near-collineation is of the form* S_1TS_2 *where* S_1 *and* S_2 *are similitudes and* T *is any one of the three transformations* e^z, $\log z$, z^n. The family preserved is a pencil of straight lines (besides the $2\infty^1$ minimal lines).

---

[1] Presented to the Society, February 24, 1940.

[2] Kasner and De Cicco, *Characterization of the Moebius group of circular transformations*, Proceedings of the National Academy of Sciences, vol. 25 (1939), pp. 209–213.

[3] In the previous paper, we derived these results for the point transformations of the *real* cartesian plane. But these same results may easily be derived for the complex cartesian plane without any difficulty. Note that a given family F of geometric configurations in the complex cartesian plane is said to possess ∞^n configurations if each of these is determined uniquely by n complex constants.

[4] Kasner, *The problem of partial geodesic representation*, Transactions of this Society, vol. 7 (1906), pp. 200–206. Also see Kasner, *The characterization of collineations*, this Bulletin, vol. 9 (1903), pp. 545–546; and Prenowitz, *The characterization of plane collineations in terms of homologous families of lines*, Transactions of this Society, vol. 38 (1935), pp. 564–599.

For our purposes, we shall find it convenient to define a point by the minimal coordinates (u, v) instead of the usual cartesian coordinates (x, y). The minimal and cartesian coordinates of the complex cartesian plane are connected by the two independent linear relations

$$u = x + iy, \qquad v = x - iy. \tag{1}$$

2. **The form of the differential equation of the two invariant families $(2\infty^1)$ of circles.** In minimal coordinates, the ∞^3 circles (excluding the ∞^2 points and the $2\infty^1$ minimal lines) of the complex plane are represented by the ∞^3 hyperbolas which possess as asymptotes the minimal lines $u=$const., and $v=$const. Thus the family of ∞^3 circles is given by the equation

$$a_0uv + a_1u + a_2v + a_3 = 0, \tag{2}$$

where a_0, a_1, a_2, a_3 are complex constants such that either $a_0(a_1a_2-a_3) \neq 0$ or $a_0=0$, $a_1a_2\neq 0$. From this equation, it follows that the differential equation of the third order of the entire family of ∞^3 circles is

$$2pp'' = 3p'^2, \qquad p = dv/du. \tag{3}$$

To obtain the set of conformal (direct and reverse) near-Moebius transformations, it is only necessary to obtain the set of *direct* conformal near-Moebius transformations. For any reverse conformal near-Moebius transformation is the product of a direct conformal near-Moebius transformation by a reflection through the x axis (the straight line $u+v=0$). Hence, in what follows, we shall only consider the set of direct conformal near-Moebius transformations.

In minimal coordinates, any direct conformal transformation is given by

$$U = \phi(u), \qquad V = \psi(v), \qquad \phi_u\psi_v \neq 0. \tag{4}$$

Upon extending this conformal transformation three times, we obtain

$$\begin{aligned}
P &= \frac{\psi_v}{\phi_u}\,p,\\
P' &= \frac{\psi_v}{\phi_u^2}\,p' + \frac{\psi_{vv}}{\phi_u^2}\,p^2 - \frac{\psi_v\phi_{uu}}{\phi_u^3}\,p,\\
P'' &= \frac{\psi_v}{\phi_u^3}\,p'' + \frac{3\psi_{vv}}{\phi_u^3}\,pp' - \frac{3\psi_v\phi_{uu}}{\phi_u^4}\,p' + \frac{\psi_{vvv}}{\phi_u^3}\,p^3 - \frac{3\psi_{vv}\phi_{uu}}{\phi_u^4}\,p^2\\
&\quad + \left(-\frac{\psi_v\phi_{uuu}}{\phi_u^4} + \frac{3\psi_v\phi_{uu}^2}{\phi_u^5}\right)p.
\end{aligned} \tag{5}$$

For those circles which become circles under our conformal transformation, we know that the differential condition (3) must be preserved. Upon applying these conditions to our conformal transformation, we obtain the following theorem.

THEOREM 1. *The only possible circles which become circles under the conformal transformation* (4) (*not of the Moebius type*) *are the* $2\infty^1$ *circles whose differential equation is of the form*

$$p^2 = \frac{\psi_v^2(2\phi_u\phi_{uuu} - 3\phi_{uu}^2)}{\phi_u^2(2\psi_v\psi_{vvv} - 3\psi_{vv}^2)}. \tag{6}$$

We note that the two families ($2\infty^1$) of circles (if they exist) are orthogonal.

3. **The $2\infty^1$ circles whose differential equation is of the form** (6). The two families of circles are given by a differential equation of the form

$$p = \pm \frac{\gamma(u)}{\delta(v)}. \tag{7}$$

We shall find all families of circles whose differential equation is of this form. By means of the Moebius transformations, we shall reduce our results to canonical forms.

For the time being we shall consider only the plus sign. We observe first that *neither* γ *nor* δ *can be zero.* For then the circles would be the minimal lines u or $v=$const. These are excluded from consideration.

Upon taking the first and second derivatives of (7), we obtain

$$\begin{aligned} p' &= \frac{\gamma_u}{\delta} - \frac{\gamma^2\delta_v}{\delta^3}, \\ p'' &= \frac{\gamma_{uu}}{\delta} - \frac{3\gamma\gamma_u\delta_v}{\delta^3} - \frac{\gamma^3\delta_{vv}}{\delta^4} + \frac{3\gamma^3\delta_v^2}{\delta^5}. \end{aligned} \tag{8}$$

Substituting (7) and (8) into (3), we find that (7) *represents the differential equation of* $2\infty^1$ *circles if and only if*

$$\frac{2\gamma\gamma_{uu} - 3\gamma_u^2}{\gamma^4} = \frac{2\delta\delta_{vv} - 3\delta_v^2}{\delta^4}. \tag{9}$$

This equation will be true if and only if each side is equal to the same complex constant a^2. Hence

$$2\gamma\gamma_{uu} - 3\gamma_u^2 = a^2\gamma^4, \qquad 2\delta\delta_{vv} - 3\delta_v^2 = a^2\delta^4. \tag{10}$$

We proceed to discuss the solution of these two differential equations. We shall divide this discussion into several parts.

Case A. *Let* $a = 0$. *In this case, the equation* (7) *represents two orthogonal parabolic pencils of circles* (*or their Moebius equivalents*). By a Moebius transformation, these may be reduced to the two orthogonal parallel pencils of straight lines

$$v = \pm u + \text{const.} \tag{A'}$$

Hence the canonical form of the differential equation (7) for Case A is

$$p^2 = 1. \tag{A}$$

(I) First let neither γ nor δ be constants. Then (10) may be written in the form

$$2\frac{d}{du}\left(\frac{\gamma_u}{\gamma}\right) = \left(\frac{\gamma_u}{\gamma}\right)^2, \qquad 2\frac{d}{dv}\left(\frac{\delta_v}{\delta}\right) = \left(\frac{\delta_v}{\delta}\right)^2. \tag{11}$$

The integration of these yields

$$\gamma = \frac{b}{(u - u_0)^2}, \qquad \delta = \frac{c}{(v - v_0)^2}, \tag{12}$$

where u_0, v_0, b, c are constants. Substituting these into (7) and integrating the resulting differential equation, we find

$$\frac{c}{v - v_0} = \pm\frac{b}{u - u_0} + \text{const.} \tag{13}$$

These are two orthogonal parabolic pencils of circles. By an appropriate Moebius transformation, we can convert (13) into (A'). Hence the canonical form of our differential equation (7) for this case is (A).

(II) Next let γ and δ be constants. By integrating (7), we find that the two families are two orthogonal parallel pencils of straight lines (which are equivalent by a Moebius transformation to two orthogonal parabolic pencils of circles). By a similitude, we can carry these into (A'). Hence the canonical form of (7) in this case is also equation (A).

(III) Let γ be not constant and δ constant. We find that γ and δ are given by

$$\gamma = \frac{-b}{(u - u_0)^2}, \qquad \delta = \text{const.}, \tag{14}$$

where u_0 and b are constants. Substituting these into (7) and integrating, we obtain

$$\delta v = \pm \frac{b}{u - u_0} + \text{const.} \tag{15}$$

The first family consists of all circles of the same radius $r = (b/\delta)^{1/2}$ and with centers on the minimal line $u = u_0$. The orthogonal family consists of all circles with centers on the minimal line $u = u_0$ and of radius ir. These two families of circles have the same radical axis, namely, their common line of centers $u = u_0$. Also these circles are tangent (and orthogonal) to each other at the point at infinity on this minimal line $u = u_0$.

These two families are equivalent by a Moebius transformation to two orthogonal parabolic pencils of circles. By an appropriate Moebius transformation, these two families (15) can be carried into the two orthogonal parallel pencils of straight lines (A′). Hence the canonical form of (7) in this case is also (A).

(IV) Finally let γ be constant and δ not constant. Then γ and δ are given by

$$\gamma = \text{const.}, \qquad \delta = -\frac{c}{(v - v_0)^2}. \tag{16}$$

Substituting these into (7) and integrating the resulting differential equation, we find

$$\frac{c}{v - v_0} = \pm \gamma u + \text{const.}, \tag{17}$$

where v_0 and c are constants.

The first family consists of all circles of the same radius $r = (c/\gamma)^{1/2}$ and with centers on the minimal line $v = v_0$. The orthogonal family consists of all circles with centers on the minimal line $v = v_0$ and of radius ir. The two families of circles have the same radical axis, namely, their common line of centers $v = v_0$. Also these circles are tangent (and orthogonal) to each other at the point at infinity on this minimal line $v = v_0$.

These two families are equivalent by a Moebius transformation to two orthogonal parabolic pencils of circles. By an appropriate Moebius transformation, these two families (17) can be converted into (A′). Thus the canonical form of (7) in this case is again (A).

Case B. *Let* $a \neq 0$. *In this case, the differential equation* (7) *represents two orthogonal nonparabolic pencils of circles* (*or their Moebius equiva-*

lents). By a Moebius transformation, these may be reduced to the pencil of straight lines with vertex at origin and to the pencil of concentric circles with their common center at the origin

$$v/u = \text{const.}, \qquad uv = \text{const.} \tag{B'}$$

Hence the canonical form of the differential equation (7) for this case is

$$p^2 = v^2/u^2. \tag{B}$$

In this case, γ (and δ) cannot be constant. Let $\lambda = \gamma\gamma_u$. The first of equations (10) can be written in the form

$$2\gamma\lambda\lambda_\gamma - 5\lambda^2 = a^2\gamma^6. \tag{18}$$

Multiplying this by γ^{-6} and integrating, we find

$$\lambda^2 = \gamma^5(a^2\gamma + 2b), \tag{19}$$

where a and b are constants. Replacing λ by $\gamma\gamma_u$, this equation (and a similar equation for δ) can be written in the form

$$\gamma_u = \gamma^2(a^2 + 2b/\gamma)^{1/2}, \qquad \delta_v = \delta^2(a^2 + 2c/\delta)^{1/2}, \tag{20}$$

where a, b, c are constants.

(I) Let neither b nor c be zero. The preceding equations then yield

$$\gamma = \frac{2b}{b^2(u - u_0)^2 - a^2}, \qquad \delta = \frac{2c}{c^2(v - v_0)^2 - a^2}. \tag{21}$$

Substituting these into (7) and integrating, we obtain the two families of circles

$$\begin{aligned} \frac{c(v - v_0) - a}{c(v - v_0) + a} &= \text{const.} \frac{b(u - u_0) - a}{b(u - u_0) + a}, \\ \frac{c(v - v_0) - a}{c(v - v_0) + a} &= \text{const.} \frac{b(u - u_0) + a}{b(u - u_0) - a}. \end{aligned} \tag{22}$$

These two families represent two orthogonal nonparabolic pencils of circles. The circles of the first family are those passing through the two points $(u_0+a/b, v_0+a/c)$ and $(u_0-a/b, v_0-a/c)$. The circles of the second family are those passing through the two points $(u_0+a/b, v_0-a/c)$ and $(u_0-a/b, v_0+a/c)$.

The two orthogonal nonparabolic pencils of circles (22) have the same center (u_0, v_0). This is the intersection of the two orthogonal radical axes and of the two orthogonal lines of centers. If d is the

distance between the two fixed points of the first pencil of circles, then *id* is the distance between the two fixed points of the second pencil. These properties are characteristic for two orthogonal nonparabolic pencils of circles.

By an appropriate Moebius transformation, the two families (22) can be converted into (B′). Thus the canonical form of the differential equation (7) for this case is (B).

(II) Let $b=c=0$. From (20), we find

$$\gamma = \frac{1}{a(u-u_0)}, \qquad \delta = \frac{1}{a(v-v_0)}. \tag{23}$$

Substituting these into (7) and integrating, we obtain the two families of circles

$$\frac{v-v_0}{u-u_0} = \text{const.}, \qquad (u-u_0)(v-v_0) = \text{const.} \tag{24}$$

The first family is a pencil of straight lines with vertex at (u_0, v_0). The second family is a pencil of concentric circles with the common center at (u_0, v_0). By an appropriate translation, these two families can be carried into (B′). Thus, in this case, the canonical form of the differential equation (7) is (B).

(III) Let $b=0$ and $c\neq 0$. In this case, we have

$$\gamma = \frac{1}{a(u-u_0)}, \qquad \delta = \frac{2c}{c^2(v-v_0)^2 - a^2}. \tag{25}$$

Substituting these into (7) and integrating, we find the two orthogonal pencils of circles

$$\begin{aligned} \frac{c(v-v_0)-a}{c(v-v_0)+a} &= \text{const.}\ (u-u_0), \\ \frac{c(v-v_0)-a}{c(v-v_0)+a} &= \text{const.}\ \frac{1}{u-u_0}. \end{aligned} \tag{26}$$

The circles of the first family are those passing through the point $(u_0, v_0+a/c)$ and the point at infinity on the minimal line $v=v_0-a/c$. The circles of the second family are those passing through the point $(u_0, v_0-a/c)$ and the point at infinity on the minimal line $v=v_0+a/c$.

The radical axis and the line of centers (or the line of centers and the radical axis) of the first (or second) family of circles are respectively the minimal lines $v=v_0+a/c$ and $v=v_0-a/c$.

Under the group of Moebius transformations, these two families are equivalent to any two orthogonal nonparabolic pencils of circles. By an appropriate Moebius transformation, these two families can be converted into (B′). Thus the canonical form of the differential equation (7) for this case is (B).

(IV) Let $b \neq 0$ and $c = 0$. The two families of circles are then

$$(27)\quad v - v_0 = \text{const.}\frac{b(u-u_0)-a}{b(u-u_0)+a}, \quad v - v_0 = \text{const.}\frac{b(u-u_0)+a}{b(u-u_0)-a}.$$

The circles of the first family are those passing through the point $(u_0+a/b, v_0)$ and the point at infinity on the minimal line $u = u_0 - a/b$. The circles of the second family are those passing through the point $(u_0-a/b, v_0)$ and the point at infinity on the minimal line $u = u_0+a/b$.

The radical axis and the line of centers (or the line of centers and the radical axis) of the first (or the second) family of circles are respectively the minimal lines $u = u_0+a/b$ and $u = u_0-a/b$.

Under the group of Moebius transformations, these two families are equivalent to any two orthogonal nonparabolic pencils of circles. By an appropriate Moebius transformation, these two families can be converted into (B′). Thus the canonical form of the differential equation (7) for this case is (B).

Thus in all cases, we have proved the following theorem:

THEOREM 2. *The only* $2\infty^1$ *circles whose differential equation is of the form* (7) *are two orthogonal pencils of circles. Under the group of Moebius transformations, these may be classified into two distinct types*: (1) *two orthogonal parabolic pencils of circles and* (2) *two orthogonal nonparabolic pencils of circles.*

The canonical forms of the differential equation (7) of types (1) and (2) are respectively (A) and (B).

4. **The conformal near-Moebius transformations.** We now proceed to find the conformal near-Moebius transformations. First it is seen that *the inverse* N^{-1} *of any conformal near-Moebius transformation* N *is also a conformal near-Moebius transformation.* For since N is a one-to-one correspondence which carries $2\infty^1$ circles into circles, it follows that the transformed circles must be $2\infty^1$ in number. Hence the inverse N^{-1} carries $2\infty^1$ circles into circles, and therefore it must be a conformal near-Moebius transformation.

Let N be any conformal near-Moebius transformation. Then there exist exactly $2\infty^1$ circles which are preserved by N. Denote these circles by γ and their transformed circles under N by Γ. That is,

$N(\gamma)=\Gamma$, and $\gamma=N^{-1}(\Gamma)$. By Theorems 1 and 2, we find that γ *and* Γ *are each two orthogonal pencils of circles.*

Let γ_0 denote the canonical form (A′) and (A), or (B′) and (B) of two orthogonal parabolic, or nonparabolic, pencils of circles. Let M_1 by any Moebius transformation which carries the two orthogonal pencils of circles γ into the canonical form γ_0, and let M_2 be any Moebius transformation which converts the canonical form γ_0 into the two orthogonal pencils of circles Γ. Then the transformation $T=M_2^{-1}NM_1^{-1}$ preserves the canonical form γ_0 of two orthogonal pencils of circles. Hence *any conformal near-Moebius transformation* N *is of the form* M_2TM_1 *where* M_1 *and* M_2 *are Moebius transformations and* T *is any conformal near-Moebius transformation which preserves the canonical form* γ_0 *of two orthogonal pencils of circles.*

Next we shall show that any transformation T which preserves the canonical form γ_0 of two orthogonal pencils of circles must be one of the three types: (1) $U=be^{au}$, $V=ce^{\pm av}$; (2) $U=a\log u+b$, $V=\pm a\log v+c$; and (3) $U=bu^n$, $V=cv^{\pm n}$. Obviously T converts either (I) (A) into (A), or (II) (A) into (B), or (III) (B) into (A), or (IV) (B) into (B).

(I) Let T convert (A) into (A). The differential equation $p^2=1$ must be preserved. For this to be so, we find that T must be of the form $U=au+b$, $V=\pm av+c$ where $a\neq 0$, b, c are constants. This of course is a simple case of the Moebius transformations. Hence we conclude that *there are no conformal near-Moebius transformations which preserve two orthogonal parabolic pencils of circles.*

(II) Let T convert (A) into (B). The differential equation $p^2=1$ is converted into $P^2=V^2/U^2$. For this to be so, we find that T must be of the form $U=be^{au}$, $V=ce^{\pm av}$, where a, b, c are nonzero constants. Hence we conclude that *any conformal near-Moebius transformation* N *which carries two orthogonal parabolic pencils of circles into two orthogonal nonparabolic pencils of circles must be of the form* M_2TM_1 *where* M_1 *and* M_2 *are Moebius transformations and* T *is the transformation* $U=e^u$, $V=e^v$.

(III) Let T convert (B) into (A). The differential equation $p^2=v^2/u^2$ is converted into $P^2=1$. Hence T must be of the form $U=a\log u+b$, $V=\pm a\log v+c$, where $a\neq 0$, b, c are constants. Thus we conclude that *any conformal near-Moebius transformation* N *which carries two orthogonal nonparabolic pencils of circles into two orthogonal parabolic pencils must be of the form* M_2TM_1 *where* M_1 *and* M_2 *are Moebius transformations and* T *is the transformation* $U=\log u$, $V=\log v$.

(IV) Let T convert (B) into (B). The differential equation $p^2=v^2/u^2$ must be carried into the differential equation $P^2=V^2/U^2$.

Hence T must be of the form $U=bu^n$, $V=cv^{\pm n}$ where n, b, c are nonzero constants. Thus *any conformal near-Moebius transformation N which carries two orthogonal nonparabolic pencils of circles into two orthogonal nonparabolic pencils of circles must be of the form M_2TM_1 where M_1 and M_2 are Moebius transformations and T is the transformation $U=u^n$, $V=v^n$ with n a complex nonzero constant.*

We therefore obtain from the preceding results the following:

THEOREM 3. *Any conformal near-Moebius transformation N of the complex cartesian plane is of the form M_2TM_1 where M_1 and M_2 are Moebius transformations and T is any of the three transformations* (1) $U=e^u$, $V=e^v$; (2) $U=\log u$, $V=\log v$; and (3) $U=u^n$, $V=v^n$ (*n a complex nonzero constant*).

In this paper, we have given all the conformal near-Moebius transformations, that is, all the conformal transformatons which preserve *exactly* $2\infty^1$ circles. In a later paper, we shall give the set of all nonconformal near-Moebius transformations, that is, the set of all nonconformal transformations which preserve *exactly* $2\infty^2$ circles.

COLUMBIA UNIVERSITY AND
BROOKLYN COLLEGE

A NOTE ON THE DEFINITION OF ARC-SETS[1]

W. L. AYRES

The interesting subsets of a Peano space which were later called the arc-sets or A-sets were defined independently and in a different manner by G. T. Whyburn[2] and the author,[3] and each studied the properties of these sets which have proved so useful in the later theory. The similarity of the properties led shortly to the conjecture that a relation existed between the two definitions, and investigation proved that they were equivalent. These are the properties (1) and (3) of the theorem of the present note. Two more equivalent properties, the properties (2) and (4) of the present note, were observed immediately.[4]

The present note introduces two new properties and shows that all six are equivalent, so that any one may be taken as the definition of the arc-sets or A-sets.

THEOREM. *For a nondegenerate*[5] *subset A of a Peano space P the following six properties are equivalent:*

(1) *There is a set H such that A is the set of points of all arcs of P whose end points belong to H.*

(2) *A contains every arc of P whose end points belong to A.*

(3) *A is a connected collection of cyclic elements of P.*

(4) *For any connected subset C of P, the set $A \cdot C$ is connected.*

(5) *Every set separating*[6] *two points of A in A separates them in P.*

(6) *A is connected, $\overline{A} \cdot \overline{C}$ is a single point for each component C of $P-A$, and this point y belongs to A if $\overline{A}$ contains two continua S and T such that $S \cdot T = y$.*

---

[1] Presented to the Society, December 1, 1939.

[2] *Concerning the structure of a continuous curve*, American Journal of Mathematics, vol. 50 (1928), pp. 167–194.

[3] *Concerning the arc-curves and basic sets of a continuous curve*, Transactions of this Society, vol. 30 (1928), pp. 567–578.

[4] Properties (1), (3), and (4) were proved equivalent by W. L. Ayres, *Concerning the arc-curves and basic sets of a continuous curve, second paper*, Transactions of this Society, vol. 31 (1929), pp. 595–612. That (1), (2), (3), and (4) are equivalent was remarked with an indication of proof by C. Kuratowski and G. T. Whyburn, *Sur les éléments cycliques et leurs applications*, Fundamenta Mathematicae, vol. 16 (1930), pp. 305–331. See page 321.

[5] Even in the trivial degenerate case, (1), (2), (4), (5), and (6) are still equivalent if we define a point x as being an arc from x to x.

[6] We use the word *separating* here in the weak or *coupure* sense, that is, a set K *separates* two points x and y in a set A if $x+y \subset A$ and there exists no connected set C such that $x+y \subset C \subset A-K$.

(1)→(2). This has been proved by the author.[7]

(2)→(3). It follows directly from (2) that A is connected, in fact, arc-wise connected. Let E be a cyclic element of P such that $A \cdot E \neq 0$. If $A \cdot E$ is a single point, this point is a cut point of P and thus is itself a cyclic element. Suppose $A \cdot E \supset x+z$ and let y be any other point of E. Then E contains an arc[8] xyz which belongs to A by property (2). Hence E belongs to A.

(3)→(4). This has been proved by Kuratowski and Whyburn.[9]

(4)→(5). Let K denote a set separating x and y in A. If there exists a connected set C such that $x+y \subset C \subset P-K$, then $C \cdot A$ is a connected subset of $A-K$ containing $x+y$ by (4). Then K does not separate x and y in A.

(5)→(6). The set A is connected, for otherwise the null set would separate two points of A in A, which it could not in P.

If $C=y$ is a degenerate component of $P-A$, $\overline{C} \cdot \overline{A} = y$. As A is connected, any degenerate component of $P-A$ is a non-cut point of P. Then y is an end point or belongs to some simple closed curve J of P. As y is a point component of $P-A$, J contains two points x and z of A. Let α and β denote the subarcs xy and yz of the arc xyz of J. There exists an open set U such that $U \supset \alpha - y$ and $\overline{U} \cdot \beta = y$. This open set may be obtained by summing all open spheres whose centers are points w of $\alpha - y$ and whose radii are less than the distance from w to β. Then $K = F(U) - y$ separates x and z in A but does not separate them in P as it contains no point of the arc xyz. Hence every degenerate component of $P-A$ is an end point. As y is an end point of P, there do not exist in P, and thus not in $\overline{A}$, two continua S and T such that $S \cdot T = y$.

Suppose that C is nondegenerate and $\overline{A} \cdot \overline{C} \supset x+z$. As C is connected, $C-x-z \neq 0$. We shall show that P contains an arc xyz, where $y \subset C$. If there is a point y in P which separates x and z, then the connectivity of C gives $C \supset y$ and any arc xz of P contains y. If no point of P separates x and z, there is a true cyclic element $E \supset x+z$. As each component of $P-E$ is an open set and has just one limit point in E, it follows that $P-E+x+z$ contains no connected subset containing $x+z$. But as $C+x+z$ is such a connected set, C must contain a point y of E distinct from x and from z. Then E contains[8] an arc xyz.

As $x+z \subset \overline{A} \cdot \overline{C}$, each neighborhood of x and z respectively contains points x' and z' of A. If these neighborhoods are chosen sufficiently

[7] *Arc-curves*, Theorem 3.

[8] W. L. Ayres, *Continuous curves which are cyclicly connected*, Bulletin de l'Académie Polonaise des Sciences et des Lettres, 1928, pp. 127–142, Theorem 3.

[9] Loc. cit., Theorem 8.5.

small, there exist small arcs xx' and zz' such that the set $xx'+xyz+zz'$ contains an arc $x'yz'$. Then just as above we may construct an open set U such that $F(U)-y$ separates x' and z' in A but not in P. Thus $\overline{A}\cdot\overline{C}$ is a single point for every component C of $P-A$.

Suppose continua S and T exist such that $S+T\subset\overline{A}$ and $S\cdot T=y$. Suppose $y\notin A$. Then S contains a point x of A, for otherwise S belongs to a component C_1 of $P-A$ and $\overline{A}\cdot\overline{C}_1\supset S$, which is not a single point. Similarly T contains a point z of A. By the same process as outlined above, there exists an open set U such that $U\supset S-y$, $\overline{U}\cdot T=y$. Then $F(U)-y$ separates x and z in A but not in P as it contains no point of the connected $S+T$.

(6)$\rightarrow$(1). Let $H=A$. Then the set of points of all arcs of P whose end points belong to H certainly contains A. Suppose there is an arc xyz such that $x+z\subset H=A$ and $y\notin A$. Let C denote the component of $P-A$ containing y. The subarc xy contains a point x of A and a point y of C, and thus contains a point of $\overline{A}\cdot\overline{C}$. Similarly the subarc yz contains a point of $\overline{A}\cdot\overline{C}$. But $\overline{A}\cdot\overline{C}$ is a single point. Hence $\overline{A}\cdot\overline{C}=y$.

As y was any interior point of the arc which belonged to $P-A$ and we showed that $\overline{A}\supset y$, we have $\overline{A}\supset$ arc xyz. Let $S=$subarc xy and $T=$subarc yz. We have $S+T\subset\overline{A}$, $S\cdot T=y$, and hence $y\subset A$ by property (6).

THE UNIVERSITY OF MICHIGAN

ON REARRANGEMENTS OF SERIES[1]

RALPH PALMER AGNEW

1. **Introduction.** Let E denote the metric space in which a point x is a permutation $x_1, x_2, x_3, \cdots$ of the positive integers and the distance (x, y) between two points $x \equiv \{x_1, x_2, \cdots\}$ and $y \equiv \{y_1, y_2, \cdots\}$ of E is given by the Fréchet formula

$$(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|x_n - y_n|}{1 + |x_n - y_n|}.$$

The space E is of the second category (Theorem 2).

Let $c_1 + c_2 + \cdots$ be a convergent series of real terms for which $\sum |c_n| = \infty$. To simplify typography, we write $c(n)$ for c_n. To each $x \in E$ corresponds a rearrangement $c(x_1) + c(x_2) + \cdots$ of the series $\sum c_n$. By a well known theorem of Riemann, $x \in E$ exists such that $c(x_1) + c(x_2) + \cdots$ converges to a preassigned number, diverges to $+\infty$ or to $-\infty$, or oscillates in a preassigned fashion.

The set A of $x \in E$ for which $c(x_1) + c(x_2) + \cdots$ converges is therefore a proper subset of E, and M. Kac has proposed the problem of determining whether $E - A$ is of the second category. The following theorem shows not only that A is of the first category (and hence that $E - A$ is of the second category) but also that the set of $x \in E$ for which the series $c(x_1) + c(x_2) + \cdots$ has unilaterally bounded partial sums is of the first category.

THEOREM 1. *For each $x \in E$ except those belonging to a set of the first category,*

$$\liminf_{N \to \infty} \sum_{n=1}^{N} c(x_n) = -\infty, \qquad \limsup_{N \to \infty} \sum_{n=1}^{N} c(x_n) = \infty.$$

2. **Proof of Theorem 1.** The fact that the "coordinates" x_n and y_n of two points x and y of E are integers implies roughly that, if N is large, then $x_n = y_n$ for $n = 1, 2, \cdots, N$ if and only if (x, y) is near 0. To make this precise, let $x \in E$, $r > 0$, and let $S(x, r)$ denote the set of points y such that $(x, y) < r$, so that $S(x, r)$ is an open sphere with center at x and radius r. It is easy to show that if x and y are two points of E such that $y \in S(x, 2^{-N-1})$ then $x_n = y_n$ when $n = 1, 2, \cdots, N$; and that if x and y are such that $x_n = y_n$ when $n = 1, 2, \cdots, N$ then $y \in S(x, 2^{-N})$.

[1] Presented to the Society, October 28, 1939.

To prove Theorem 1, let B denote the set of $x \in E$ for which

$$\limsup_{N\to\infty} \sum_{n=1}^{N} c(x_n) < \infty;$$

and, for each $h>0$, let B_h denote the set of $x \in E$ for which

$$\text{l.u.b.}_{N=1,2,3,\cdots} \sum_{n=1}^{N} c(x_n) < h.$$

Then

$$B = B_1 + B_2 + B_3 + \cdots.$$

We show that B is the first category by showing that B_h is nondense for each $h>0$. Suppose $h>0$ exists such that the closure $\overline{B}_h$ of B_h contains a sphere S' with center at $x' \equiv \{x_1', x_2', \cdots\}$ and radius $r>0$. Choose m so great that $2^{-m-1}+2^{-m-2}+\cdots<r/2$. Let $x_n''=x_n'$ when $1 \leqq n \leqq m$; and define x_n'' for $n>m$ in such a way that $\sum c(x_n'')$ diverges to $+\infty$. Then $(x', x'')<r/2$ so that $x'' \in S'$. Choose an index q such that

$$c(x_1'') + c(x_2'') + \cdots + c(x_q'') > h,$$

and then choose $\delta>0$ such that $x_k=y_k$ for $k=1, 2, \cdots, q$ whenever $x, y \in E$ and $(x, y)<\delta$.

If x is a point within the sphere S'' with center at x'' and radius δ (that is, if $(x, x'')<\delta$), then $c(x_1)+c(x_2)+\cdots+c(x_q)>h$ and x is not in B_h. Thus B_h contains no point of S'' and consequently $\overline{B}_h$ does not contain x''. This contradicts the assumption that $\overline{B}_h$ contains S', and hence proves that B_h is nondense and B is of the first category. Similar considerations show that the set C of $x \in E$ for which $c(x_1)+\cdots+c(x_N)$ has inferior limit greater than $-\infty$ is of the first category. Since the union of two sets B and C of the first category is itself of the first category, Theorem 1 is established.

If $z_1+z_2+\cdots$ is a convergent series of complex terms for which $\sum|z_n|=\infty$, it is easy to apply our theorem to the series of real and imaginary parts of z_n to show that the set of $x \in E$ for which $z(x_1)+z(x_2)+\cdots$ has bounded partial sums is a set of the first category.

3. **The space E.** In this section we obtain some properties of E and prove the following result.

THEOREM 2. *The space E is of the second category at each of its points.*

That the space E is not complete was pointed out to the author by Professor L. M. Graves. In fact if $x^{(n)}$ is the point

$$x^{(n)} \equiv \{2, 3, \cdots, n-1, n, 1, n+1, n+2, \cdots\}$$

of E, then $x^{(n)}$ is a Cauchy sequence in E which does not converge to a point of E. If $\mathcal{E}$ is the space in which a point is a sequence of positive integers not necessarily a permutation of all positive integers, and the distance between two points of $\mathcal{E}$ is given by the Fréchet formula, then $\mathcal{E}$ is complete and E is a subspace of $\mathcal{E}$. It is easy to show that the closure of E in $\mathcal{E}$ is the space $\mathcal{E}_1$ in which a point is a sequence of positive integers containing each positive integer *at most once*, and hence that $\mathcal{E}_1$ is the least complete subspace of $\mathcal{E}$ which contains E. For example, $\{2, 4, 6, 8, \cdots\}$ is a point of $\mathcal{E}_1$ which is not a point of E.

If $\mathcal{E}_x\{x_n = k\}$ denotes, for each n, $k = 1, 2, \cdots$, the set of all $x \in \mathcal{E}$ for which $x_n = k$, then

$$\mathcal{E}_2 \equiv \prod_{k=1}^{\infty} \sum_{n=1}^{\infty} \mathcal{E}_x\{x_n = k\}$$

is the subset of $\mathcal{E}$ in which a point is a sequence containing each positive integer *at least once*. Since $\mathcal{E}_x\{x_n = k\}$ is an open subset of $\mathcal{E}$ for each n, $k = 1, 2, \cdots$, $\mathcal{E}_2$ is the intersection of a countable set of open sets (that is, $\mathcal{E}_2$ is a G_δ) in $\mathcal{E}$. Since $\mathcal{E}_1$ is a closed subset of $\mathcal{E}$ and $E = \mathcal{E}_1\mathcal{E}_2$, it follows that E is a G_δ in the complete space $\mathcal{E}$.

Therefore, by a fundamental theorem whose proof is an easy extension of the familiar proof that a complete metric space is of the second category, E is of the second category at each of its points and Theorem 2 is proved.

CORNELL UNIVERSITY

ABELIAN GROUPS THAT ARE DIRECT SUMMANDS OF EVERY CONTAINING ABELIAN GROUP[1]

REINHOLD BAER

It is a well known theorem that an abelian group G satisfying $G=nG$ for every positive integer n is a direct summand of every abelian group H which contains G as a subgroup. It is the object of this note to generalize this theorem to abelian groups admitting a ring of operators, and to show that the corresponding conditions are not only sufficient but are at the same time necessary. Finally we show that every abelian group admitting a ring of operators may be imbedded in a group with the above mentioned properties; and it is possible to choose this "completion" of the given group in such a way that it is isomorphic to a subgroup of every other completion.

Our investigation is concerned with abelian groups admitting a ring of operators. A ring R is an abelian group with regard to addition, its multiplication is associative, and the two operations are connected by the distributive laws. As the multiplication in R need not be commutative, we ought to distinguish left-, right- and two-sided ideals. Since, however, only left-ideals will occur in the future, we may use the term "ideals" without fear of confusion. Thus an ideal in R is a non-vacuous set M of elements in R with the property:

If m', m'' are elements in M, and if r', r'' are elements in R, then $r'm'\pm r''m''$ is an element in M.

An abelian group G whose composition is written as addition is said to admit the elements in the ring R as operators (or shorter: G is an *abelian group over* R), if with every element r in R and g in G is connected their uniquely determined product rg so that this product is an element in G and so that this multiplication satisfies the associative and distributive laws. If G is an abelian group over R, then its subgroups M are characterized by the same property as the ideals M in R.

We assume finally the existence of an element 1 in R so that $g=1g$ for every g in G and $r\cdot 1=1\cdot r=r$ for every r in R.

If x is any element in the abelian group G over R, then its *order* $N(x)$ consists of all the elements r in R which satisfy $rx=0$. One verifies that every order $N(x)$ is an ideal in R, and that $N(x)=R$ if, and only if, $x=0$.

If M is an ideal in R, and if x is an element in G, then a subgroup

[1] Presented to the Society, February 24, 1940.

of G is formed by the elements mx for m in M; and this subgroup may be denoted by Mx. (It is a subgroup of the cyclic group generated by x.) The correspondence between the element m in M and the element mx in Mx is a special case of the homomorphisms of M into G. Here a *homomorphism* ϕ of the ideal M in R into the abelian group G over R is defined as a single-valued function m^ϕ of the elements m in M with values in G which satisfies

$$(r'm' \pm r''m'')^\phi = r'(m'^\phi) \pm r''(m''^\phi)$$

for m', m'' in M and r', r'' in R.

We are now ready to state and prove our main result.[2]

THEOREM 1. *The following two properties of an abelian group G over the ring R are each a consequence of the other.*

(a) *If G is a subgroup of the abelian group H over R, then G is a direct summand of H.*

(b) *To every ideal M in R and to every homomorphism ϕ of M into G there exists some element v in G so that $m^\phi = mv$ for every m in M.*

PROOF. Assume first that (a) is satisfied by G. If M is an ideal in R, and if ϕ is a homomorphism of M into G, then there exists one and essentially only one group H over R which is generated in adjoining to G an element h, subject to the relations

$$mh = m^\phi \qquad \text{for every } m \text{ in } M.$$

It is a consequence of (a) that H is the direct sum of G and of a suitable subgroup K of H so that every element in H may be represented in one and only one way in the form: $g+k$ for g in G and k in K. This applies in particular to the element h so that $h=v+w$ for

---

[2] The following is a remark by the referee: "It is perhaps of some interest to observe that Theorem 1 contains a generalization of the theorem that every representation of a semisimple algebra is fully reducible. Indeed, how does one characterize those rings R such that *every* abelian group G admitting R as an operator ring has the property (a)? The answer is that every left-ideal in R must be generated by an idempotent element, and this is equivalent to saying that R is semisimple (both chain conditions and no radical).

"The sufficiency of this condition is proved by showing that G has the property (b). Let $m \to m^\phi$ be any homomorphism of a left-ideal M of R into G. Since $M=Re$ with $e^2=e$, then $me=m$ for every m in M, and if we take $v=e^\phi$ we have $m^\phi=(me)^\phi=me^\phi=mv$.

"The necessity is proved by observing that every left-ideal M of R is itself an abelian group over R, and is a subgroup of R. Hence to each M there exists a complementary left-ideal N. From $1=e+e'$ with unique e in M, e' in N, one concludes in the usual way that $e^2=e$ and $M=Re$."

uniquely determined elements v and w in G and K respectively. Then $m^\phi = mh = mv + mw$ is for every element m in M an element in G, and so is mv and $m^\phi - mv = mw$. Since mw is an element in K, it follows that $mw = 0$ or $m^\phi = mv$ for every m in M; and this shows that (b) is a consequence of (a).

Assume now conversely that (b) is satisfied by the abelian group G over R, and that G is a subgroup of the abelian group H over R. Then there exists a greatest subgroup K of H whose meet with G is 0. The subgroup S of H which is generated by G and K is their direct sum; and hence it suffices to prove that $H = S = G + K$.

If $S \neq H$, then there exists in H an element w that is not contained in S. The coset $W = S + w$ is an element in the quotient-group H/S and its order $N(W)$ is an ideal in R. If m is any element in $N(W)$, then mw is an element in S; and it follows from the construction of S that there exist uniquely determined elements $g(m)$ and $k(m)$ in G and K respectively so that

$$mw = g(m) + k(m) \qquad \text{for } m \text{ in } N(W).$$

Thus a homomorphism of $N(W)$ into G is defined in mapping the element m in $N(W)$ upon the element $g(m)$ in G. There exists therefore by condition (b) an element v in G so that $mv = g(m)$ for every m in $N(W)$. The element $w' = w - v$ consequently satisfies

$$S + w' = S + w = W$$

and

$$mw' = k(m)$$

for every element m in $N(W)$. Since w is not an element in S, neither is w'. Since K is a greatest subgroup of H whose meet with G is 0, and since w' is not an element in S and therefore not in K, adjoining w' to K generates a subgroup whose meet with G is different from 0. Hence there exists an element k in K, and an element r in R so that

$$k + rw' = g \neq 0$$

is an element in G. Since $rw' = g - k$ is an element in S, it follows that r is an element in $N(W)$ so that $g = k + k(r)$ is an element in the meet of G and K. This contradicts, however, the construction of K and the choice of g. Our hypothesis $S \neq H$ has thus led us to a contradiction; and this completes the proof.

If M is an ideal in the ring R, and if G is an abelian group over R, then G is termed *M-complete*, if there exists to every homomorphism ϕ of M into G an element v in G so that $m^\phi = mv$ for every m in M. In

this terminology, condition (b) of Theorem 1 states that G is M-complete for every ideal M in R. It is our object to characterize the M-complete groups, provided M is a principal ideal. For this end we need several notations.

If M is an ideal in R, then denote by G_M the set of all the elements g in G which satisfy $mg=0$ for every m in M. Note that G_M need not be a subgroup of the abelian group G over R, though it is closed with regard to addition and subtraction.

If p is any element in R, then pG consists of all the elements pg for g in G. Note again that pG need not be a subgroup of the abelian group G over R.

The principal ideal in R, generated by the element p in R, consists of all the elements rp for r in R and may therefore be denoted by Rp; and $N(p)$ consists of all the elements r in R so that $rp=0$. $N(p)$ is clearly an ideal in R.

THEOREM 2. *The abelian group G over R is Rp-complete if and only if $G_{N(p)} \leqq pG$.*

PROOF. Suppose first that G is Rp-complete. If x is an element in $G_{N(p)}$, then $r'p=r''p$ implies $r'x=r''x$, since the first equation is equivalent to the fact that $r'-r''$ is an element in $N(p)$. Thus a homomorphism of Rp into G is defined in mapping rp upon rx. Since G is Rp-complete, there exists an element v in G so that $rpv=rx$ for every r in R. This implies in particular that $pv=x$, that is, our condition is necessary.

Suppose conversely that our condition be satisfied by G. If ϕ is a homomorphism of Rp into G, then p^ϕ is an element in $G_{N(p)}$, since we have $r(p^\phi)=(rp)^\phi=0$ for elements r in $N(p)$. Hence there exists an element q in G so that $p^\phi=pq$; and clearly $(rp)^\phi=r(p^\phi)=r(pq)=(rp)q$ for every r in R so that G is Rp-complete.

COROLLARY 1. *If p is an element in R so that $N(p)=0$, then $G=pG$ is a necessary and sufficient condition for Rp-completeness of the abelian group G over R.*

This is a consequence of Theorem 2, since $G_0=G$.

COROLLARY 2. *If $N(p)=0$ for every element $p\neq 0$ in the ring R, and if every ideal in R is a principal ideal Rp, then $G=pG$ for every $p\neq 0$ in R is a necessary and sufficient condition for the abelian group G over R to be a direct summand of every abelian group H over R which contains G as a subgroup.*

This is an obvious consequence of Theorem 1 and Corollary 1.

If, in particular, R consists of the rational integers, then the hypotheses of Corollary 2 are satisfied. In this case the sufficiency of the condition of the Corollary 2 has been known for a long time.[3]

An abelian group G over the ring R is termed *complete*, if it is M-complete for every ideal M in R. Thus the complete groups are just the groups satisfying the properties (a) and (b) of Theorem 1.

THEOREM 3. *Every abelian group over the ring R is a subgroup of a complete abelian group over the ring R.*

PROOF. If G is an abelian group over the ring R, M an ideal in R, and ϕ a homomorphism of M into G, then there exists an abelian group H over R which contains G as a subgroup and which contains an element x so that $mx=m^{\phi}$ for every m in M.

By repetition of the construction of the preceding paragraph one may show that if G is an abelian group over the ring R, then there exists an abelian group G' over the ring R which contains G as a subgroup and which satisfies the following condition:

(3.1) If M is an ideal in R, and if ϕ is a homomorphism of M into G, then there exists an element v in G' so that $mv=m^{\phi}$ for every m in M.

Denote now by λ an ordinal number which is a limit-ordinal and whose cardinal number is greater than the number of elements in R. Then it follows from the second paragraph of the proof that there exists for every ordinal ν with $0\leqq\nu\leqq\lambda$ an abelian group G_{ν} over R with the following properties:

(i) $G_0=G$;

(ii) $G_{\nu}\leqq G_{\mu}$ for $\nu<\mu$;

(iii) G_{ν} is for limit-ordinals ν the set of all the elements contained in groups G_{μ} for $\mu<\nu$;

(iv) G_{ν} and $G_{\nu+1}$ satisfy condition (3.1).

Suppose now that M is an ideal in R and that ϕ is a homomorphism of M into $H=G_{\lambda}$. Then there exists an ordinal $\sigma<\lambda$ so that G_{σ} contains all the elements m^{ϕ}; and there exists therefore an element v in $G_{\sigma+1}$ so that $mv=m^{\phi}$ for every m. H is therefore complete.

THEOREM 4. *To every subgroup G of the complete abelian group K over the ring R there exists a complete subgroup G^* of K which contains G as a subgroup and which satisfies the following condition:*

(E) *Every isomorphism of G upon a subgroup of a complete abelian group H over R is induced by an isomorphism of G^* upon a subgroup of H.*

---

[3] For a comparable proof see Annals of Mathematics, (2), vol. 37 (1936), pp. 766–767, (1; 1).

PROOF. If T is an abelian group over the ring R, if M is an ideal in R, then the homomorphism ϕ of M into T is termed *reducible* in T, if there exists an ideal M' in R and a homomorphism ϕ' of M' into T so that $M<M'$ and so that ϕ and ϕ' coincide on M. If ϕ is not reducible, then it is irreducible in T.

(4.1) *The abelian group T over the ring R is complete, if there exists to every ideal M in R and to every irreducible homomorphism ϕ of M into T an element v in T so that $mv=m^{\phi}$ for every m in M.*

To prove this statement let J be an ideal in R and γ a homomorphism of J into T. Then there exists a greatest ideal M in R so that $J\leqq M$ and so that γ is induced in J by a homomorphism ϕ of M into T. It is clear that ϕ is irreducible in T. Hence there exists an element v in T so that $mv=m^{\phi}$ for every m in M. This implies however that $nv=n^{\phi}=n^{\gamma}$ for every n in J, that is, T is complete.

It is a consequence of (4.1) and of the completeness of K that there exists an ascending chain of subgroups G_ν for $0\leqq\nu\leqq\lambda$ with the following properties:

(i) $G=G_0$;

(ii) $G_\nu\leqq K$ for $\nu\leqq\lambda$;

(iii) $G_{\nu+1}$ is generated by adjoining to G_ν an element g_ν with the following properties:

(iii′) The homomorphism of $N(G_\nu+g_\nu)$ into G_ν which is defined by mapping the element m in $N(G_\nu+g_\nu)$ upon the element mg_ν in G_ν is irreducible in G_ν.

(iii″)[4] G_ν does not contain any element x so that $mx=mg_\nu$ for every m in $N(G_\nu+g_\nu)$.

(iv) G_ν is for limit-ordinals ν the set of all the elements contained in groups G_μ for $\mu<\nu$.

(v) $G_\lambda=G^*$ is complete.

We are now going to prove that this subgroup G^* of K satisfies condition (E). Thus assume that ρ is an isomorphism of G upon the subgroup $G'=G^{\rho}$ of the complete group H. We are going to construct subgroups G'_ν of H and isomorphisms ρ_ν of G_ν upon G'_ν with the following properties:

(1) $G'=G'_0$, $\rho=\rho_0$;

(2) $G'_\nu\leqq G'_\mu$ for $\nu\leqq\mu$;

(3) ρ_ν and ρ_μ coincide on G_ν for $\nu\leqq\mu$.

In order to prove the possibility of this construction it suffices to show the existence of $G'_{\nu+1}$, $\rho_{\nu+1}$ under the hypothesis of the existence of G'_ν, ρ_ν.

[4] This condition (iii″) is not really needed for the proof, though it is convenient for the construction of the chain G_ν.

A homomorphism irreducible in G_ν' of $M=N(G_\nu+g_\nu)$ into G_ν' is defined by mapping the element m in M upon the element $m^\phi=(mg_\nu)\rho_\nu$. Since H is complete, there exists an element h in H so that $m^\phi=mh$ for every m in M. If $M'=N(G_\nu'+h)$, then it is clear that $M\leqq M'$. If m is in M', then mh is an element in G_ν'. Thus a homomorphism γ of M' into G_ν is defined by mapping the element m in M' upon the element $m^\gamma=(mh)\rho_\nu^{-1}$. If, in particular, m is an element in M, then $m^\gamma=mg_\nu$; and it follows from (iii') that $M=M'$. Suppose now that g' is an element in G_ν' and u an element in R so that $g'+uh=0$. Then u is an element in $M=M'$ and it follows from the above considerations that $-g'=uh=u^{\gamma\rho_\nu}=(ug_\nu)\rho_\nu$. Hence there exists one and only one isomorphism $\rho_{\nu+1}$ of $G_{\nu+1}$ upon the group $G_{\nu+1}'$, generated by G_ν' and h, which isomorphism induces ρ_ν in G_ν and maps g_ν upon h.

Thus there exists finally an isomorphism ρ_λ of $G^*=G_\lambda$ upon G_λ' which induces ρ in G; and this completes the proof.

COROLLARY. *Assume that K is a smallest complete abelian group over the ring R containing the subgroups G_i. Then G_1 and G_2 are isomorphic if, and only if, there exists an automorphism of K mapping G_1 upon G_2.*

This is an obvious consequence of Theorem 4. It should be noted, however, that the complete group G^*, satisfying (E) and containing G, whose existence is assured by Theorems 3 and 4, is only "essentially smallest," but need not be "actually smallest."

UNIVERSITY OF ILLINOIS

CLOSURE OF PRODUCTS OF FUNCTIONS[1]

D. G. BOURGIN

This note presents some natural theorems on the characterizations of certain closed (*or complete*) sets of functions with separable variables. In order to motivate the developments of the paper we treat a simple case first in elaborate detail. The proof is so formulated that it holds with trifling modifications for the more general situations in Theorems 3 and 4. The result in Theorem 5 belongs to a slightly different range of ideas.

Let $s\sim(s_1, \cdots, s_m)$ and $t\sim(t_1, \cdots, t_n)$ here stand for points in the euclidean spaces R_m and R_n. The term "interval" designates the generalized rectangular parallelopipedon open on the left.[2] We shall make use of the intervals $I_s \subset R_m$, $I_t \subset R_n$ and $I_2 = I_s \times I_t \subset R_{n+m}$. We are first interested in $L_2(I)$, the space of complex valued functions of summable square over I. The norm and scalar product are defined as usual by

$$(1) \qquad \|f(s,t) - g(s,t)\| = \left[\int_{I_t}\int_{I_s} |f(s,t) - g(s,t)|^2 dI_s dI_t\right]^{1/2},$$

$$(2) \qquad (f(s,t), g(s,t)) = \int_{I_t}\int_{I_s} f(s,t)\bar{g}(s,t) dI_s dI_t,$$

where $\bar{g}(s,t)$ is the conjugate of $g(s,t)$. The subscript I_s or I_t will indicate that the left-hand functionals are on the corresponding intervals.

We shall understand closure of the sequence of functions[3] $\{\phi_\gamma(t)\psi_\mu(s)\}$, $\gamma, \mu = 0, 1, \cdots$, to mean that for every $f(s,t) \in L_2(I_2)$ and arbitrary $\epsilon > 0$ there exists a finite sequence of complex constants $\{\beta_{\gamma\mu}\}$ and integers A and B such that

$$(3) \qquad \left\| f(s,t) - \sum_0^A \sum_0^B \beta_{\gamma\mu}\phi_\gamma(t)\psi_\mu(s) \right\| < \epsilon.$$

It is well known that with the adjunction of the scalar product defined in (2), $L_2(I_2)$ is a complex Hilbert space and that closure and completeness are equivalent concepts.

THEOREM 1. *If* $\{\phi_\gamma(t)\psi_\mu(s)\}$, $\gamma, \mu = 0, 1, \cdots$, *is a sequence of com-*

[1] Presented to the Society, December 2, 1939.

[2] S. Saks, *Theory of the Integral*, English edition, p. 57.

[3] Curly brackets, $\{\}$, will always denote sequences.

plex valued functions in $L_2(I_2)$, then a necessary and sufficient[4] *condition for closure is that $\{\phi_\gamma(t)\}$ and $\{\psi_\mu(s)\}$ be closed in the spaces $L_2(I_t)$ and $L_2(I_s)$ respectively.*

We deal with the sufficiency demonstration first. Suppose the denumerable set of all subintervals, with rational end points, of I_t to be ordered according to 0, 1, 2, $\cdots$. We designate by $h_\rho(t)$ the characteristic function[5] of the ρth subinterval divided by its norm. The function $g_\nu(s)$ is similarly defined for the range I_s. Thus

$$\|h_\rho(t)\|_{I_t} = \|g_\nu(s)\|_{I_s} = 1. \tag{4}$$

It is well known that $\{h_\rho(t)g_\nu(s)\}$, $\rho, \nu = 0, 1, \cdots$, has the closure property in $L_2(I_2)$. Hence for $f(s, t) \in L_2(I_2)$ and arbitrary $\epsilon > 0$ we can find integers M and N and MN complex constants $\{a_{\rho\nu}\}$ such that

$$\left\| f(s, t) - \sum_0^M \sum_0^N a_{\rho\nu} h_\rho(t) g_\nu(s) \right\| < \epsilon/2. \tag{5}$$

Let

$$\delta \leqq \min\left(\frac{\epsilon}{4MN} \max |a_{\rho\nu}|, 1\right). \tag{6}$$

Thus

$$2\delta \sum_0^M \sum_0^N |a_{\rho\nu}| < \epsilon/2. \tag{6.1}$$

In view of the assumed closure properties of $\{\phi_\gamma(t)\}$ and $\{\psi_\mu(s)\}$, integers A and B and complex constants $\{d_\mu^{(\nu)}\}$, $\{e_\gamma^{(\rho)}\}$, $\rho = 0, 1, \cdots, M$ and $\nu = 0, 1, \cdots, N$, exist which yield the simultaneous inequalities

$$\left\| g_\nu(s) - \sum_{\mu=0}^B d_\mu^{(\nu)} \psi_\mu(s) \right\|_I < \delta/2, \tag{7}$$

$$\left\| h_\rho(t) - \sum_0^A e_\gamma^{(\rho)} \phi_\gamma(t) \right\|_{I_t} < \delta/2. \tag{7.1}$$

Hence

[4] A special case amounting to the assertion of sufficiency, only, for the subspace of $L_2(I_2)$ composed of real continuous functions, when $\{\phi_\gamma(t)\}$ and $\{\psi_\mu(s)\}$ are restricted to be orthogonal sets of functions, has been given by Courant: Courant-Hilbert, *Methoden der mathematischen Physik*, vol. 1, 1st edition, p. 90. Another special sufficiency proof is given in A. Zymund, *Trigonometrical Series*, p. 13.

[5] Saks, loc. cit., p. 6.

$$(7.2)\qquad \left\| \sum_0^B d_\mu^{(\nu)}\psi_\mu(s) \right\|_{I_s} \leqq \|g_\nu(s)\|_{I_t} + \left\| g_\nu(s) - \sum_0^B d_\mu^{(\nu)}\psi_\mu(s) \right\|_{I_s} \leqq 2.$$

Let $\beta_{\gamma\mu} = \sum_{\rho=0}^{M}\sum_{\nu=0}^{N} a_{\rho\nu} e_\gamma^{(\rho)} d_\mu^{(\nu)}$. The triangle inequality for norms yields, in view of (6), (7), (7.11), and (7.2)

$$\begin{aligned}
&\left\| h_\rho(t)g_\nu(s) - \sum_0^A\sum_0^B e_\gamma^{(\rho)} d_\mu^{(\nu)} \phi_\gamma(t)\psi_\mu(s) \right\| \\
&\leqq \left\| h_\rho(t)\left(g_\nu(s) - \sum_0^B d_\mu^{(\nu)}\psi_\mu(s) \right) \right\| \\
&\quad + \left\| \sum_0^B d_\mu^{(\nu)}\psi_\mu(s)\left(h_\rho(t) - \sum_0^A e_\gamma^{(\rho)}\phi_\gamma(t) \right) \right\| \qquad (8) \\
&\leqq \|h_\rho(t)\|_{I_t} \left\| g_\nu(s) - \sum_0^B d_\mu^{(\nu)}\psi_\mu(s) \right\|_{I_s} \\
&\quad + \left\| \sum_0^B d_\mu^{(\nu)}\psi_\mu(s) \right\|_{I_s} \left\| h_\rho(t) - \sum_0^A e_\gamma^{(\rho)}\phi_\gamma(t) \right\|_{I_t} \\
&\leqq 2\delta, \qquad \text{for } \rho = 0, 1, \cdots, M,\ \nu = 0, 1, \cdots, N.
\end{aligned}$$

On combining the various inequalities above

$$\begin{aligned}
&\left\| f(s, t) - \sum_0^A\sum_0^B \beta_{\gamma\mu}\phi_\gamma(t)\psi_\mu(s) \right\| \\
&\leqq \left\| f(s, t) - \sum_0^M\sum_0^N a_{\rho\nu}h_\rho(t)g_\nu(s) \right\| \\
&\quad + \left\| \sum_0^M\sum_0^N a_{\rho\nu}\left(h_\rho(t)g_\nu(s) - \sum_0^A\sum_0^B e_\gamma^{\rho} d_\mu^{\nu}\phi_\gamma(t)\psi_\mu(s) \right) \right\| \qquad (9) \\
&\leqq \epsilon/2 + \sum_0^M\sum_0^N \left(|a_{\rho\nu}| \left\| h_\rho(t)g_\nu(s) - \sum_0^A\sum_0^B e_\gamma^{(\rho)} d_\mu^{(\nu)}\phi_\gamma(t)\psi_\mu(s) \right\| \right) \\
&\leqq \epsilon/2 + 2\delta\sum_0^M\sum_0^N |a_{\rho\nu}| \leqq \epsilon.
\end{aligned}$$

This asserts the closure property for $\{\phi_\gamma(t)\psi_\mu(s)\}$.

The necessity demonstration is equally direct. A trivial application of Fubini's theorem indicates that $\phi_\gamma(t) \in L_2(I_t)$, $\psi_\mu(s) \in L_2(I_s)$ when $\phi_\gamma(t)\psi_\mu(s) \in L_2(I_2)$. No generality is lost if we assume that $\{\psi_\mu(s)\}$ is a linearly independent set of functions. Suppose $\{\psi_\mu(s)\}$ does not have the closure property. Then $f(s) \in L_2(I_s)$ exists for which for all R and b_μ

$$\text{(10)} \qquad \text{G.L.B.} \left\| f(s) - \sum_0^R b_\mu \psi_\mu(s) \right\|_{I_s} = c > 0, \qquad b_\mu = b'_\mu + ib''_\mu.$$

A fundamental result of Riesz guarantees the existence of *minimal* constants,[6] $\{b_\mu^R\}$, such that for $b_\mu \not\equiv b_\mu^R$, $\mu \leqq R$,

$$\text{(11)} \qquad \left\| f(s) - \sum_0^R b_\mu^R \psi_\mu(s) \right\|_{I_s} \leqq \left\| f(s) - \sum_0^R b_\mu \psi_\mu(s) \right\|_{I_s}.$$

The corresponding minimal constants for $Af(s)$ are evidently $\{Ab_\mu^R\}$. Hence[7]

$$\text{(12)} \qquad \left\| F(t)f(s) - \sum_0^R b_\mu^R F(t)\psi_\mu(s) \right\|_{I_s} \leqq \left\| F(t)f(s) - \sum_0^R b_\mu(t)\psi_\mu(s) \right\|_{I_s}, \quad t \in I_t,$$

when $F(t) \in L_2(I_t)$ is a fixed function of positive norm. We write

$$\text{(13)} \qquad b_\mu(t) = \sum_0^Q a_{\gamma\mu}\phi_\gamma(t), \qquad Q < \infty.$$

In view of (12) we have

$$\text{(14)} \qquad \begin{aligned} 0 < c\|F(t)\|_{I_t} &\leqq \left\| f(s)F(t) - \sum_0^R b_\mu^R F(t)\psi_\mu(s) \right\| \\ &= \left[\int_{I_t} \left\| f(s)F(t) - \sum_0^R b_\mu^R F(t)\psi_\mu(s) \right\|_{I_s}^2 dI_t \right]^{1/2} \\ &\leqq \left[\int_{I_t} \left\| f(s)F(t) - \sum_0^R \sum_0^Q a_{\gamma\mu}\phi_\gamma(t)\psi_\mu(s) \right\|_{I_s}^2 dI_t \right]^{1/2} \\ &= \left\| f(s)F(t) - \sum_0^R \sum_0^Q a_{\gamma\mu}\phi_\gamma(t)\psi_\mu(s) \right\|. \end{aligned}$$

Since (14) is in contradiction with the assumed closure property of $\{\phi_\gamma(t)\psi_\mu(s)\}$ our necessity proof is complete.

We denote by $h'_\rho(t)$ and $g'_\nu(s)$ the step functions in R_n and R_m analogous to $h_\rho(t)$ and $g_\nu(s)$. According to a classical result, $\{h'_\rho(t)g'_\nu(s)\}$, $\rho, \nu = 0, 1, \cdots$, have the closure property in $L_2(E_2)$ when the s, t integration is over R_{n+m} or any Lebesgue measurable subset E_2. Accordingly Theorem 1 and its demonstration remain formally valid in detail when I_s, I_t and I_2 are replaced either by R_n, R_m and R_{n+m} or by

[6] F. Riesz, Acta Mathematica, vol. 41 (1916), p. 77, Lemma 3.

[7] With the choice $F(t)f(s)$, the method of proof of the necessity condition remains valid when sets of infinite measure are included.

the sets E_s, E_t and $E_2 = E_s \times E_t$ of finite or infinite Lebesgue measure.

THEOREM 2. *If* $\{F_\rho(s, t)\}$, $\rho = 0, 1, \cdots$, *is closed in* $L_2(E_2)$, *then the sequence is also closed in* $L_2(E_s)$ *except possibly for a* t *set of zero measure.*[8]

Suppose a lower bound of approximation to $f(s) \in L_2(E_s)$, by linear combinations of $\{F_\rho(s, t)\}$, is $c(t) \in L_2(E_t)$, where $\infty > c(t) > 0$ for $t \in G \subset E_t$. Let $F(t) \in L_2(E_t)$ differ from 0 on G (say $F(t) = c(t)$). The analogue of (14) is

$$(14') \qquad \|c(t)F(t)\|_G \leqq \|c(t)F(t)\|_{E_t} \leqq \left\| f(s)F(t) - \sum_0^R b_\mu F_\mu(s, t) \right\|.$$

Hence G has zero measure. Let $\{f^{(\sigma)}(s)\}$ be closed in $L_2(E_s)$ and denote the corresponding G sets, defined above, by $\{G^\sigma\}$. The denumerable sum $\mathfrak{S}G^\sigma$ is plainly of measure zero. Thus $\{F_\rho(s, t)\}$ is closed in $L_2(E_s)$ for all $t \in E_t - \mathfrak{S}G^\sigma$.

We now abstract the properties needed in the foregoing proofs. Let $T(E)$ denote a Banach space[9] of real functions on E. A set G, $G \subset E$, will be called a *non-significant* set if $f(z) \in T(E)$ may be arbitrarily changed on G without affecting the value of $\|f(z)\|_E$. The postulates below hold for $T(E)$. When (d) and (e) are omitted we write $T_-(E)$.

(a) If $f(s, t) \in T(E_2)$ then $f(s, t) \in T(E_s)$ and $f(s, t) \in T(E_t)$ for all save a non-significant set of t or s values respectively. If $f(s) \in T(E_s)$, $F(t) \in T(E_t)$ then $f(s)F(t) \in T(E_2)$.

(b) $\|f(s, t)\|_{E_2} = \| \|f(s, t)\|_{E_s} \|_{E_t}$.

(c) If, neglecting non-significant sets, $|f_1(t)| > |f_2(t)|$, then $\|f_1(t)\|_{E_t} > \|f_2(t)\|_{E_t}$.

(d) There exists a sequence $\{h_\rho(t)g_\nu(s)\}$, $\rho, \nu = 0, 1, \cdots$, with the closure property in $T(E_2)$, where $h_\rho(t) \in T(E_t)$ and $g_\nu(s) \in T(E_s)$.

(e) Denumerable sums of non-significant sets are non-significant sets.

---

[8] A sharper result follows from Fatou's lemma. Suppose $F(t) \in L_2(E_t)$ differs from 0 for almost all $t \in E_t$. Now

$$0 = \mathrm{L}_{N\to\infty} \left\| f(s)F(t) - \sum_0^N b_\rho^{(N)} F_\rho(s, t) \right\|_{E_2}^2 \geqq \int_{E_t} \mathrm{L}_{N\to\infty} \left\| f(s)F(t) - \sum_0^N b_\rho^{(N)} F_\rho(s, t) \right\|_{E_s}^2 dE_t.$$

Thus a suitable sequence $\{\sum_N b_\rho^{(i)} F_\rho(s, t)\}$, with *constant coefficients* $\{b_\rho^{(i)}\}$, converges strongly to $f(s)$ in $L_2(E_s)$ for almost all $t \in E_t$. Moreover if E_t is of finite measure, the Egoroff theorem guarantees uniform convergence for $t \in D_\delta \subset E_t$ where the measure of $E_t - D_\delta$ is inferior to arbitrary δ. A closed sequence $\{f_\sigma(s)\}$ is introduced as above.

[9] S. Banach, *Théorie des Opérations Linéaires*, pp. 53, 58. Banach uses *fundamental* in the sense of our *closed*.

THEOREM 3. (α) *If* $\{\phi_\gamma(t)\}$ *and* $\{\psi_\sigma(s)\}$ *are closed in* $T(E_1)$, $T(E_2)$ *then* $\{\phi_\gamma(t)\psi_\sigma(s)\}$ *is closed in* $T(E_2)$. (β) *If* $\{\phi_\gamma(t)\psi_\sigma(s)\}$ *is closed in* $T_-(E_2)$, *then* $\{\psi_\sigma(s)\}$ *is closed in* $T_-(E_2)$. (γ) *If* $\{F_\rho(s, t)\}$ *is closed in* $T(E_2)$, *then* $\{F_\rho(s, t)\}$ *is closed in* $T(E_2)$ *for all but a non-significant set of* t *values in* E_1.

The demonstrations of Theorems 1 and 2 apply without change in form.[10] The space[11] $L_p(E, \mu)$, $p \geqq 1$, is included in $T(E)$. This is the space of measurable functions whose pth powers are summable over the measurable set E, where the Lebesgue-Radon-Stieltjes integral is equally admissible with the usual Lebesgue integral. Thus the symbol $\mu(E)$ denotes either the Lebesgue measure, or the Radon measure determined by a non-negative additive function of intervals. In all cases $\mu_2(E_2) = \mu_\sigma(E_\sigma)\mu_t(E_t)$, and the sets of zero measure constitute the non-significant sets. The norm is

$$\|f(s, t)\| = \left[\int_E |f|^p d\mu(E)\right]^{1/p}. \tag{15}$$

The verification of the main postulates is implied by the Fubini theorem, the Hölder-Minkowski inequalities and the denseness of the step functions. The functions $\{h_\rho(t)\}$, $\{g_\sigma(s)\}$ or $\{h'_\rho(t)g'_\sigma(s)\}$ as defined in Theorem 1 are again available.[12]

The space $C(E)$ of continuous functions is another special case of $T(E)$. We assume $E_t \subseteqq R_m$, $E_t \subseteqq R_n$ and $E_2 \subseteqq R_{n+m}$ are bounded closed sets. The null set is the only non-significant set. The norm is

$$\|f(s, t)\| = \max_{s,\, t \in E_2} |f(s, t)|. \tag{16}$$

The sequences $h_\rho(t)$ and $g_\sigma(s)$ are the ordered products of the elements $1, t_1, \cdots, t_n$ and of $1, s_1, \cdots, s_m$ respectively.

Postulates (b) and (c) may be replaced by the weaker

(b′) $\|f(w, z)\|_{E_z} < \epsilon$ implies $\|f(w, z)\|_{E_w} < \eta(\epsilon)$, where $\lim_{\epsilon \to 0} \eta(\epsilon) = 0$ except possibly for non-significant z sets.

(c′) $\|G(z)\|_{E_z} = 1$, $\|H(w)\|_{E_w} < \epsilon$ imply $\|G(z)H(w)\|_{E_z} < \eta(\epsilon)$.

[10] For (α), postulate (d) may be replaced by the assumption that each $f(s, t) \in T(E_2)$ is the strong limit of some (not necessarily fixed) sequence $\{h'_\rho(t)g'_\sigma(s)\}$, where $h'_\rho(t) \in T(E_t)$ and $g'_\sigma(s) \in T(E_s)$.

[11] Saks, loc. cit. (1928), chap. 3, or J. Radon, Sitzungsberichte der Akademie der Wissenschaften, Vienna, class IIa, vol. 122 (1913). The Lebesgue case admits sets of infinite measure.

[12] For $p > 1$ a valid theorem on *completeness* is obtained from Theorem 3 if closure (in $L_p(E, \mu)$) is replaced by completeness in $L_{p/(p-1)}(E, \mu)$ where E refers to E_σ, E_t and E_2 in turn.

These modifications will be connoted by writing $T'(E)$ and $T'_-(E)$. Consider, for instance, $C^1(E)$, the space of functions continuous together with their first partial derivatives on[13] E. We restrict ourselves now to *closed* linear intervals I_s, I_t and the rectangle I_2: $I_s \times I_t$. The norms in $C^1(I_2)$ and $C^1(I_s)$ are,[14] with $f_s \equiv \partial f/\partial s$,

$$\|f(s,t)\| = \max_{I_2} |f(s,t)| + \max_{I_2} |f_s(s,t)| + \max_{I_2} |f_t(s,t)|,$$
$$\|f(s)\| = \max_{I_s} |f(s)| + \max_{I_s} |f_s(s)|. \tag{17}$$

It is well known that $C^1(I_s)$ (and $C^1(I_t)$) is complete. It is easy to show that $C^1(I_2)$ also is complete. Indeed if $\{f^{(n)}(s,t)\}$ is a Cauchy sequence in $C^1(I_2)$, then $f^{(n)}(s,t)$, $f_s^{(n)}(s,t)$ and $f_t^{(n)}(s,t)$ converge *uniformly* in I_2 and hence define an element of $C^1(I_2)$.

Since

(b′) $\|F(s,t)\|_{I_2} \geqq \max_{t \in I_t} \|f(s,t)\|_{I_s}$ (t and s are interchangeable),

(c′) $\|G(s)H(t)\|_{I_2} \leqq \|G(s)\|_{I_s} \|H(t)\|_{I_t}$,

it is clear that (b′) and (c′) are satisfied.

THEOREM 4. *The conclusions in* (α), (β), (γ) *of Theorem* 3 *remain valid when* $T'(E)$ *and* $T'_-(E)$ *replace* $T(E)$ *and* $T_-(E)$.

For (α) we now choose δ small enough in (7) and (7.1) to yield $\eta(\delta)$ inferior to the right side of (6). Then (6.1) is valid with $\eta(\delta)$ written in place of δ. On making use of (c′) it is easily shown that the left side of (8) is smaller than $2\eta(\delta)$ and the final inequality in (9) is again obtained. For (β) we need only change (14) slightly. Indeed, by reference to (b′) and (13)

$$\epsilon \geqq \|f(s)F(t) - \sum\sum a_{\gamma\mu}\phi_\gamma(t)\psi_\mu(s)\|_{I_2},$$

would imply the contradiction

$$\eta(\epsilon) \geqq |c| \ \text{true max} |F(t)| > 0. \tag{14″}$$

The *true maximum* is defined just as in the analogous case of measurable functions and implies neglect of non-significant t sets. Evidently (γ) also may be maintained. Indeed the argument in footnote 8, for example, is easily amended to yield the desired result.

If the closure property of the sequence $\{\phi_\rho(z)\}$ in $T_-(E)$ or $C^1(I)$ is unaffected by the omission of $\phi_\sigma(z)$, then we shall say $\{\phi_\rho(z)\}$ is a

[13] The sets used in $C(E)$ are available for $C^1(E)$ also.

[14] Even if $f(s,t)$ and $g(s)h(t) \in C^1(E_2)$, $\|\|f(s,t)\|_{I_s}\|_{I_t}$ and $\|\|g(s)h(t)\|_{I_s}\|_{I_t}$ need not exist. Thus $C^1(E)$ is not included under $T(E)$.

"*redundant*" sequence and $\phi_\sigma(z)$ is a "*superfluous*" function. If $\{\phi_{\sigma_k}\}$, $k=1, 2, \cdots, K$, is superfluous, then for arbitrary ϵ we can satisfy

$$\left\| \phi_{\sigma_k} - \sum_0^N{}' c_j \phi_j(z) \right\| < \epsilon, \quad j \neq \sigma_l, l = 1, \cdots, K. \tag{18}$$

LEMMA 1. *In $T_-(E)$ or $T'_-(E)$ if $\{f_\mu(z)\}$ is closed and non-redundant, then for any $F(z)$, $\overline{\lim}_{\epsilon \to 0} |d_1(\epsilon)| \leqq D < \infty$ where $d_1(\epsilon)$ is consistent with $\|F(z) - d_1(\epsilon) f_1(z) - \sum_2^N d_j f_j(z)\| < \epsilon$.*

In the contrary case

$$\begin{aligned} \epsilon + \|F(z)\| &\geqq \|F(z)\| + \left\| F(z) - d_1(\epsilon) f_1(z) - \sum_2^N d_j f_j(z) \right\| \\ &\geqq |d_1(\epsilon)| \left\| f_1(z) - \sum_2^N \frac{d_i}{d_1} f_i(z) \right\|. \end{aligned} \tag{19}$$

Now

$$\left\| f_1(z) - \sum_2^N \frac{d_i}{d_1} f_i(z) \right\| \geqq c > 0, \tag{20}$$

for all N and d_i, since $f_1(z)$ is not superfluous. For all sufficiently small ϵ, (19) and (20) imply

$$|d_1(\epsilon)| \leqq 2\|F(z)\|/c \tag{21}$$

in contradiction with the hypothesized non-boundedness of $d_1(\epsilon)$.

THEOREM 5. *If $\{\phi_\mu(t)\psi_\mu(s)\}$ is closed in $T_-(E_2)$ or C^1E, then* (I) *$\{\psi_\mu(s)\}$ is closed in $T_-(E_s)$ (or $C^1(I_s)$);* (II) *every finite subsequence of $\{\psi_\mu(s)\}$ is superfluous.*[15]

Evidently (I) is a special case of Theorem 3(β). In view of (I) if $\phi_\sigma(T)\psi_\sigma(s)$, $\sigma = 1, \cdots, q$, is superfluous, then $\psi_\sigma(s)$, $\sigma = 1, \cdots, q$, is superfluous. Accordingly we may restrict ourselves to non-redundant sequences $\{\phi_\mu(t)\psi_\mu(s)\}$.

We demonstrate (II) by induction. Suppose $\psi_1(s), \cdots, \psi_{n-1}(s)$ are superfluous. Since no finite basis exists in $T_-(E)$ or $C^1(I)$, we may find a function $F(t)$ such that the set $F(t), \phi_\sigma(t)$, $\sigma = 1, \cdots, n$, is linearly independent. Suppose $\psi_n(s)$ is not superfluous. Then

$$\left\| \psi_n(s) - \sum_{n+1}^N k_i \psi_i(s) \right\|_E \geqq c > 0, \tag{22}$$

[15] Evidently $\{\psi_\mu(s)\}$ need not be *dense closed* in the sense that any infinite subsequence is closed.

for all k_i and N. By hypothesis sequences $\{a_i^{(\rho)}\}$ and a constant N exist for arbitrary ϵ such that

$$\left\| \psi_\rho(s) - \sum_{i=n}^{N} a_i^{(\rho)} \psi_i(s) \right\|_{E_s} \leqq \epsilon, \quad \rho = 1, \cdots, n-1. \tag{22.1}$$

Moreover

$$\begin{aligned} &\left\| \psi_n(s)F(t) - \sum_{\rho=1}^{n} d_\rho \phi_\rho(t)\psi_\rho(s) - \sum_{n+1}^{N} d_i\phi_i(t)\psi_i(s) \right\| \\ &\quad + \left\| \sum_{\rho=1}^{n-1} d_\rho\phi_\rho(t)\left(\psi_\rho(s) - \sum_{n}^{N} a_i^{(\rho)}\psi_i(s)\right) \right\| \\ &\quad \geqq \left\| \psi_n(s)\left[F(t) - \sum_{\rho=1}^{n-1} d_\rho a_n^\rho \phi_\rho(t) - d_n\phi_n(t) \right] \right. \\ &\quad \left. - \sum_{n+1}^{N} d_i\phi_i(t)\psi_i(s) - \sum_{\sigma=1}^{n-1}\sum_{n+1}^{N} a_i^{(\sigma)} d_\sigma\phi_\sigma(t)\psi_i(s) \right\|. \end{aligned} \tag{23}$$

The right side of this inequality, by an argument similar in all details to that involved in the passage from (12) to (14), dominates

$$c\left\| F(t) - \sum_{1}^{n-1} d_\sigma a_n^{(\sigma)} \phi_\sigma(t) - d_n\phi_n(t) \right\|_{I_t} \text{ in } T_-(E_2) \tag{23.1}$$

or (cf. (b'))

$$\max_{t \in I_t} c\left| F(t) - \sum_{1}^{n-1} d_\sigma a_n^\sigma \phi_\sigma(t) - d_n\phi_n(t) \right| \text{ in } C^1(I_2). \tag{23.2}$$

In (23.2) we note $\phi_j(t) \in C^1(I_t)$ implies $\phi_j(t) \in C(I_t)$. Hence again by the Riesz theorem the expressions in (23.1) and (23.2) have a positive lower bound, denoted by $K>0$. In view of (22.1) closure of $\{\phi_\mu(t)\psi_\mu(s)\}$ and postulates (b) or (c''), constants N, d_i and $a_i^{(\rho)}$ exist such that the left side of (23) is inferior to

$$\epsilon + \sum_{1}^{n-1} |d_\rho| \, \|\phi_\rho(t)\|_{I_t}\epsilon. \tag{24}$$

Hence by Lemma 1 applied to each d_ρ the upper bound in (24) approaches 0 with ϵ in contradiction with the conclusion $K>0$. Thus $\psi_n(s)$ is superfluous.

This type of argument may be used to show that the non-redundancy of $\{\phi_\mu(t)\psi_\mu(s)\}$ implies that $\psi_1(s)$ is superfluous. The induction is thus complete and part (II) of our theorem is established. It is an easy matter to extend the theorem to $T'_-(E)$ spaces.

UNIVERSITY OF ILLINOIS

THE SPACES L^p WITH $0<p<1$[1]

MAHLON M. DAY[2]

If we have a Lebesgue measurable set E in any n-dimensional euclidean space, and have any positive number q, we define $L^q(E)$ to be the class of all real-valued Lebesgue measurable functions f on E for which $\int_E |f|^q < \infty$. As is well known, whenever $q \geqq 1$ the class of functions $L^q(E)$ is a Banach space with the norm $\|f\|_q = (\int_E |f|^q)^{1/q}$. When $0<p<1$, the function $\|f\|_p$ no longer satisfies the triangle inequality $\|f_1+f_2\|_p \leqq \|f_1\|_p + \|f_2\|_p$ but in general only the weaker condition $\|f_1+f_2\|_p \leqq 2^\nu [\|f_1\|_p + \|f_2\|_p]$, where $\nu = (1-p)/p$. (This can be shown by considering the function $(1+x^p)/(1+x)^p$.) If we consider such an L^p space as a linear topological space in which the neighborhoods of a point f_0 are the spheres of radius $\epsilon > 0$, $E_{f \in L^p}[\|f-f_0\|_p < \epsilon]$, it follows from theorems of Hyers and Wehausen[3] that this topology can be given by an equivalent Fréchet metric. This suggests that while many theorems on Banach spaces which can be applied to the spaces $L^p(E)$ with $p \geqq 1$ may fail to hold in those spaces with $0<p<1$, there may still remain many theorems on Fréchet spaces and pseudo-normed spaces which may be applicable. However, Theorem 1 shows that almost no results depending on the use of linear (that is, additive and continuous) functionals can be usefully applied in these spaces.

THEOREM 1. *Any linear functional on $L^p(E)$ is identically zero.*

The proof of this and of some additional results is given in a series of lemmas using a more general set of assumptions. We assume as a background some knowledge of the first chapter of Saks[4] book, in which he deals with what he calls "the integral," a completely additive integral having many of the properties of the Lebesgue integral. We consider a set Y of elements y, an additive family[5] $\mathfrak{X}$ of subsets of Y and an additive, non-negative[6] set-function μ on $\mathfrak{X}$ such that $\mu(Y) < \infty$. This last condition will be imposed from here until we

[1] Presented to the Society, February 24, 1940.

[2] Corinna Borden Keen Research Fellow of Brown University.

[3] D. H. Hyers, *A note on linear topological spaces*, this Bulletin, vol. 44 (1938), pp. 76–80, and J. V. Wehausen, *Transformations in linear topological spaces*, Duke Mathematical Journal, vol. 4 (1938), pp. 157–169.

[4] S. Saks, *Theory of the Integral*, Warsaw, 1937.

[5] A class $\mathfrak{X}$ of subsets of Y is called an additive family if (a) $X \in \mathfrak{X}$ implies $Y-X \in \mathfrak{X}$, and (b) $X_n \in \mathfrak{X}$ implies $\sum_{n<\infty} X_n \in \mathfrak{X}$.

[6] A set-function μ is additive if whenever X_n are disjoint sets of $\mathfrak{X}$ then $\mu(\sum_{n<\infty} X_n) = \sum_{n<\infty} \mu(X_n)$; μ is non-negative if $\mu(X) \geqq 0$ for every $X \in \mathfrak{X}$.

reach Theorem 4. We shall finally remove it altogether and get the same type of theorem giving the linear functionals on any $L^p_\mu(Y)$. We shall also find the conjugate space of the sequence spaces related to these. For each $p>0$ let $L^p=L^p_\mu(Y)$ be the class of real-valued, $\mathfrak{X}$-measurable[7] functions f on Y such that $\|f\|_p=(\int_Y|f|^p d\mu)^{1/p}<\infty$. It is clear, as in the spaces with $p\geqq 1$, that if U is a linear functional on L^p, there is a $K\geqq 0$ such that $|U(f)|\leqq K$ whenever $\|f\|_p\leqq 1$. We let $\|U\|$ be the smallest such K; that is, $\|U\|=\sup_f|U(f)|$ where the sup is taken over those f with $\|f\|_p=1$. Similarly if an additive functional is bounded on the unit sphere it is easily seen to be continuous and hence linear.

In the series of lemmas immediately following we shall consider a fixed number p with $0<p<1$ and a fixed family $\mathfrak{X}$ and measure μ.

LEMMA 1. *If $f_i, f\in L^1$ and $\|f_i-f\|_1\to 0$, then $\|f_i-f\|_p\to 0$.*

PROOF. Let $E_i=E_{y\in Y}[|f_i(y)-f(y)|>1]$, and let $g_i(y)=f(y)$ in E_i, $g_i(y)=f_i(y)$ in $Y-E_i$. Then

$$\int_Y|f_i-f|^p d\mu=\int_{E_i}|f_i-f|^p d\mu+\int_Y|g_i-f|^p d\mu.$$

We have $\int_{E_i}|f_i-f|^p\leqq\int_{E_i}|f_i-f|\,d\mu\leqq\int_Y|f_i-f|\,d\mu$; this integral tends to zero by hypothesis. Also $\|g_i-f\|_1\to 0$, so g_i converges in measure to f; moreover, $|g_i(y)-f(y)|\leqq 1$ for all $y\in Y$, so, by the Lebesgue convergence theorem, $\int_Y|g_i-f|^p d\mu\to 0$.

LEMMA 2. *Every linear functional on L^1 can be expressed in the form $U(f)=\int_Y fu d\mu$ where u depends on U, is $\mathfrak{X}$-measurable and, except possibly on a set of μ-measure zero, is bounded.*

The proof follows that given in Banach[8] for the special case $Y=(0,1)$. If we let f_E be the characteristic function of $E\in\mathfrak{X}$ and let $\Phi(E)=U(f_E)$, the set-function Φ thus defined on $\mathfrak{X}$ is additive, $|\Phi(E)|\leqq\|U\|\mu(E)$ and $U(f)=\int_Y f d\Phi$ for each $f\in L^1$. By the theorem of Radon and Nikodym (Saks, p. 36) there is an $\mathfrak{X}$-measurable function u such that $\Phi(E)=\int_E u d\mu$ and $\|U\|=\operatorname{ess\,sup}_{y\in Y}|u(y)|$. By a theorem on change of measure (Saks, p. 37) we have $U(f)=\int_Y fu d\mu$.

LEMMA 3. *If U is linear on L^p, there is an $\mathfrak{X}$-measurable, essentially bounded function u on Y such that whenever $f\in L^1$, $U(f)=\int_Y fu d\mu$, while if $f\in L^p-L^1$ there exists a sequence $\{f_i\}$ of functions in L^1 such that*

[7] A function f defined on Y is $\mathfrak{X}$-measurable if for each real a the set $E_{y\in Y}[f(y)>a]$ is in $\mathfrak{X}$.

[8] S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932.

$\|f_i - f\|_p \to 0$, *and for every such sequence* $U(f) = \lim_{i\to\infty} U(f_i) = \lim_{i\to\infty} \int_Y f_i u d\mu$.

This follows immediately from the first two lemmas and from the obvious fact that the elements of L^1 form a dense set in L^p.

LEMMA 4. *If for each* $E \in \mathfrak{X}$ *with* $\mu(E) > 0$ *there is a function* $f_0 \in L^p - L^1$ *such that* $f_0(y) = 0$ *when* $y \in Y - E$, *and if* U *is linear on* L^p, *then* $U(f) = 0$ *for every* $f \in L^p$.

PROOF. If U is not identically zero, the function u of Lemma 3 cannot be almost everywhere zero; hence there is an $\alpha > 0$ and a set $E \in X$ such that $|u(y)| > \alpha$ if $y \in E$. Let $f'(y) = f_0(y)$ signum $f_0(y) \cdot u(y)$ where signum $s = |s|/s$ when $s \neq 0, = 0$ when $s = 0$; for each n let $E_n = E_y \in Y[|f'(y)| \leqq n]$, and let $f_n(y) = f'(y)$ on E_n, $f_n(y) = 0$ otherwise. Then $\|f_n - f'\|_p \to 0$; also

$$U(f_n) = \int_Y u f_n d\mu = \int_{E_n} |f_0 u| \, d\mu \geqq \alpha \int_{E_n} |f_0| \, d\mu,$$

but $E_n \uparrow E$ and $\int_E |f_0| \, d\mu = \infty$; so $U(f_n) \uparrow \infty$, contradicting continuity.

LEMMA 5. *If* $q > 1$, $E \in \mathfrak{X}$ *and* $\mu(E) > 0$, *and if no function* f *which is zero on* $Y - E$ *belongs to* $L^1 - L^q$, *then every* $f \in L^1$ *is essentially bounded on* E.

PROOF. If there is a function not essentially bounded on E, there is such a function f_0 which at the same time is zero on $Y - E$. Hence there exists a countably infinite sequence of sets $E_n \in \mathfrak{X}$ with $\mu(E_n) = b_n > 0$, $E_n \subset E$, E_n disjoint, $|f_0(y)| \geqq n$ on E_n, and, since $f_0 \in L^q$, $\sum n^q b_n < \infty$. Let α be so chosen that $1 < \alpha < 1/q + q - 1$ and let $a_n = 1/b_n n^\alpha$; let $f(y) = a_n$ if $y \in E_n$, $f(y) = 0$ if $y \in Y - \sum_{n<\infty} E_n$. Then $\int_Y |f| \, d\mu = \sum_{n<\infty} a_n b_n = \sum n^{-\alpha} < \infty$, so $f \in L^1$. On the other hand $\int_Y |f|^q d\mu = \sum a_n^q b_n = \sum b_n^{1-q} n^{-\alpha q}$. Since $\sum n^q b_n < \infty$ there must be a $K > 0$ such that $n^q b_n < K$ for all n; this implies that $b_n^{1-q} > n^{q(q-1)} K^{1-q}$ so $\int_Y |f|^q d\mu \geqq K^{1-q} \sum n^{q(q-1)} n^{-\alpha q} \geqq K^{1-q} \sum n^{-1} = \infty$, and $f \notin L^q$ contrary to hypothesis.

For convenience we introduce a property of sets of $\mathfrak{X}$: a set $E \in \mathfrak{X}$ has property A relative to μ (in symbols $E \in \mathfrak{A}$) if $0 < \mu(E) < \infty$ and for every $E' \in \mathfrak{X}$ either[9] $\mu(E \cdot E') = 0$ or $\mu(E - E') = 0$.

There are several obvious properties of the sets of $\mathfrak{A}$.

(1) If E has only one element and $\mu(E) > 0$, then $E \in \mathfrak{A}$.

[9] The referee has called to the writer's attention the fact that S. Saks calls such sets "singular" in his paper, *Addition to the note on some functionals*, Transactions of this Society, vol. 35 (1933), pp. 965–970.

(2) If $E \in \mathfrak{A}$ and f is $\mathfrak{X}$-measurable, then f is equal to a constant $\bar{f}$ almost everywhere on E.

(3) If $E_1 \in \mathfrak{A}$ and $E_2 \in \mathfrak{X}$, either $\mu(E_1E_2)=0$ or $\mu(E_1E_2)=\mu(E_1)$.

(4) $\mathfrak{A}$ is empty if and only if given $\epsilon>0$ it is possible to cover any set $E \in \mathfrak{X}$ with a countable collection of sets $S_i \in \mathfrak{X}$ with $\mu(S_i)<\epsilon$ for all i.

(5) If $E \in \mathfrak{A}$ and $\mu(E')=\mu(E'')=0$, then $E_0=E+E'-E'' \in \mathfrak{A}$ also.

THEOREM 2. *There is a nonzero linear functional on L^p if and only if $\mathfrak{A}$ is not empty.*

PROOF. If $\mathfrak{A}$ is empty and $E \in \mathfrak{X}$ with $\mu(E)>0$, there is a sequence of disjoint subsets of E, each of positive measure. Hence there is a function in L^1 essentially unbounded on E; therefore, by Lemma 5 with $q=1/p$, there is an $f \in L^p-L^1$ with $f(y)=0$ in $Y-E$; by Lemma 4 every linear functional on L^p is identically zero.

On the other hand if there is a set $E \in \mathfrak{A}$, and if $\|f_n-f\|_p \to 0$, f_n converges in measure to f; that is, for each $\epsilon>0$, $\mu(E_{y \in Y}[|f_n(y)-f(y)|>\epsilon]) \to 0$ as $n \to \infty$. If $\bar{f}_n$ and $\bar{f}$ are the values of f_n and f almost everywhere on E, it follows that $\bar{f}_n \to \bar{f}$. Hence for any real number k setting $U(f)=k\bar{f}$ defines a linear functional U on L^p and if $k \neq 0$ then U is not identically zero.

For the case in which $m(E)$, the Lebesgue measure of E, is finite, Theorem 1 follows directly from Theorem 2 since every Lebesgue measurable set can be split into two subsets, each of measure half as great, so $\mathfrak{A}$ is empty for Lebesgue measure; the same conclusion could be derived from (4) above. The general case follows from this and from Theorem 6 below.

There are a large number of measures μ for which $\mathfrak{A}$ is empty. For example, to get one such class of measures, let X be a separable metric space and μ a Carathéodory measure on the class $\mathfrak{X}$ of μ-measurable sets; the class $\mathfrak{X}$, as is well known, contains all the Borel sets in X. If Y is a measurable subset of X, we say that μ is a uniformly continuous function on Y if the conditions $E_n \in \mathfrak{X}$, $E_n \subset Y$, and the sequence of diameters[10] $\delta(E_n) \to 0$ together imply that $\mu(E_n) \to 0$. It is easy to show from this condition that given $\epsilon>0$ there is a $\delta>0$ such that $\mu(E_n)<\epsilon$ whenever $E \in \mathfrak{X}$, $E \subset Y$ and $\delta(E)<\epsilon$.

LEMMA 6. *If μ is a Carathéodory measure on the separable space X, if $X=\sum Y_j$ with each $Y_j \in \mathfrak{X}$ and $\mu(Y_j)<\infty$, and μ is uniformly continuous on each Y_j, then the class $\mathfrak{A}$ is empty.*

[10] $\delta(E)=\sup_{x,y}\rho(x, y)$ where the sup is taken over $x \in E$ and $y \in E$ and $\rho(x, y)$ is the distance in X from x to y.

PROOF. If $E \in \mathfrak{A}$ and $\mu(E)=b>0$, we have by (3) that $\mu(EY_i)=0$ if $i \neq j$, while $\mu(EY_j)=b$; then by the separability of X we can cover Y_j with spheres of arbitrarily small radius. Hence by uniform continuity and (4) $\mu(EY_j)=0$ also and $\mu(E)=0$, so E cannot be in $\mathfrak{A}$.

COROLLARY. *For such X and μ, if $\mu(X)<\infty$ there are no nonzero linear functionals on $L_\mu^p(X)$, $0<p<1$.*

We remark that a later theorem allows us to remove the restriction that $\mu(X)<\infty$.

The interesting spaces $L_\mu^p(Y)$, then, are those for which the set $\mathfrak{A}$ is not empty; in such a case the structure of $\mathfrak{A}$ is rather simple.

LEMMA 7. *If $\mathfrak{A}$ is not empty, there exists a finite or countable sequence of sets $\{E_i\} \subset \mathfrak{A}$ such that every $E \in \mathfrak{A}$ differs from just one E_i only by sets of measure zero.*

PROOF. Well-order $\mathfrak{A}$, and define E_α to be the first set in the ordering which does not overlap any E_β, $\beta<\alpha$, on a set of positive measure. Since $\mu(Y)<\infty$, the number of these sets E_α for which $\mu(E)>1/n$ must be finite for each n; hence the sequence $\{E_\alpha\}$ is finite or countable.

In what follows we shall let $\{E_i\}$ be the family of sets of $\mathfrak{A}$ of the preceding lemma, let $b_i=\mu(E_i)$ and let $\bar{f}_i$ be the value of the measurable function f almost everywhere on E_i.

THEOREM 3. *Any linear functional on L^p is identically zero, if $\mathfrak{A}$ is empty, or can be expressed in the form $U(f)=\sum_i u_i \bar{f}_i$ where $\|U\|=\sup_i |u_i| b_i^{-1/p}$. If u_i are given so that $|u_i|<Kb_i^{1/p}$ for all i, the functional $U(f)=\sum_i u_i f_i$ is linear on L^p.*

PROOF. It is easily seen from Theorem 2, applied to $L_\mu^p(Y-\sum_i E_i)$, that $U(f)$ is independent of the values of f on $Y-\sum_i E_i$; hence no generality is lost if we assume $Y=\sum_i E_i$ for simplicity. Let $f_{(i)}$ be the characteristic function of E_i (that is, $f_{(i)}(y)=1$ if $y \in E_i$, $f_{(i)}(y)=0$ otherwise), and take any $f \in L^p$; then $\|\sum_{i \leq n} \bar{f}_i f_{(i)} - f\|_p \to 0$ as $n \to \infty$; so, by continuity, $U(f)=\lim_n U(\sum_{i \leq n} \bar{f}_i f_{(i)})=\lim_n \sum_{i \leq n} \bar{f}_i U(f_{(i)})=\sum_i u_i \bar{f}_i$ where $u_i=U(f_{(i)})$. Postponing the computation of $\|U\|$ for a moment, let us assume u_i given such that $|u_i| \leq Kb_i^{1/p}$, and take $\|f\|_p=1$. Then $|\bar{f}_i b_i^{1/p}| \leq \|f\|_p \leq 1$, so that $|\bar{f}_i| b_i^{1/p} \leq |\bar{f}_i|^p b_i$ since $0<p<1$. Therefore $|U(f)| \leq \sum_i |u_i \bar{f}_i| \leq K \sum_i |b_i^{1/p} \bar{f}_i| \leq K \sum_i |\bar{f}_i|^p b_i \leq K$, so U is bounded; hence it is continuous, and $\|U\| \leq K$. It follows, if $K=\sup_i |u_i| b_i^{-1/p}$, that $\|U\| \geq K$ also for $|U(b_i^{-1/p} f_{(i)})| = |u_i| b_i^{-1/p} \leq \|U\|$; therefore $\sup_i |u_i| b_i^{-1/p} \leq \|U\|$ also. We note that this shows that the series $\sum_i |u_i \bar{f}_i|$ converges.

We now turn to the case in which $\mu(Y)$ need not be finite; for greater convenience we introduce another property of sets relative to μ; Y has property B (in symbols $Y \in \mathfrak{B}$) if Y can be expressed as $\sum_{j<\omega} Y_j$ with $Y_j \in \mathfrak{X}$ and $\mu(Y_j) < \infty$ for each j.

Clearly Lemma 7 holds if $Y \in \mathfrak{B}$ as well as if $\mu(Y) < \infty$; we let E_i, b_i and $\bar{f}_i$ have their previous meanings.

THEOREM 4. *If $Y \in \mathfrak{B}$, a functional U on $L_\mu^p(Y)$ is linear if and only if* (a) *it is identically zero and $\mathfrak{A}$ is empty, or* (b) *$\mathfrak{A}$ is not empty and U can be expressed in the form $U(f) = \sum_i u_i \bar{f}_i$ with $\|U\| = \sup_i |u_i| b_i^{-1/p}$.*

PROOF. The sets Y_j which exist by property B can obviously be taken disjoint; we let U_j be a linear functional on $L_\mu^p(Y)$ defined by $U_j(f) = U(f_j')$ where $f_j'(y) = f(y)$ when $y \in Y_j$, $f_j'(y) = 0$ otherwise. Then U_j is linear on $L_\mu^p(Y_j)$; so by Theorems 2 and 3 either U_j is identically zero or $U_j(f) = \sum_i u_{ji} \bar{f}_{ji}$. Now $\lim_n \|\sum_{i \leq n} f_i' - f\|_p = 0$, so $U(f) = \lim_n \sum_{j \leq n} U(f_j') = \sum_j U_j(f)$. Moreover there is an $f_0 \in L^p$ such that $|f_0(y)| = |f(y)|$ for all $y \in Y$ and $U_j(f_0) = |U_j(f)|$ for all j; hence $|U_j(f)| \leq \sum_i U_j(f_0) = U(f_0) \leq \|U\| \cdot \|f_0\| = \|U\| \cdot \|f\|$, so $\|U_j\| \leq \|U\|$ and the series $\sum_j U_j(f)$ converges absolutely for all $f \in L_\mu^p$. Then $|u_{ji}| b_{ji}^{1/p} \leq \|U_j\| \leq \|U\|$ and $U(f) = \sum_j \sum_i u_{ji} \bar{f}_{ji}$ unless all U_j are identically zero, that is, unless $\mathfrak{A} = 0$. Rearranging this into a simple series, permissible since it converges absolutely, we get $U(f) = \sum u_i f_i$. The other conclusions follow as in Theorem 3.

The general form of Theorem 1 follows directly from this, since the whole euclidean space is the sum of a countable set of finite intervals.

There remains the case in which Y is not in $\mathfrak{B}$. As we did in the proof of Lemma 7, we can define a well-ordered set $\{E_\gamma\}$ of sets of $\mathfrak{A}$, disjoint up to sets of measure zero and such that every $E \in \mathfrak{A}$ is, except for sets of measure zero, equal to some E_γ; however we have no assurance that the sequence will be countable. As before we let $b_\gamma = \mu(E_\gamma)$ and $\bar{f}_\gamma$ be the value of f almost everywhere on E_γ. We need also a known lemma about the set in which a function in $L_\mu^p(Y)$ is not equal to 0.

LEMMA 8. *For every $q > 0$ and any $f \in L_\mu^q(Y)$, $E_0 = E_y[f(y) \neq 0] \in \mathfrak{B}$.*

PROOF. Since $f \in L^q$ means $|f|^q \in L^1$, we need only consider $L_\mu^1(Y)$ and may suppose that $f(y) > 0$ for all $y \in Y$. By definition of the integral (see Saks, p. 19 ff.) $\int_Y f d\mu = \sup_g \int_Y g d\mu$ where the sup is taken over those functions g which are linear combinations of characteristic functions of sets of $\mathfrak{X}$ and which at the same time satisfy the inequalities $0 \leq g(y) \leq f(y)$ for almost all $y \in Y$. But such a function can be different from zero only on a set of finite measure if it is to have a

finite integral. There is a sequence $\{g_i\}$ of these functions such that $g_i \uparrow f$ almost everywhere in Y and $\int_Y f d\mu = \lim_i \int_Y g_i d\mu$; hence the set where $f>0$ is, except for a set of measure zero, the sum of the sets on which the $g_i>0$, and each of these sets is of finite measure.

By this lemma it is possible to put each function f of $L_\mu^p(Y)$ into at least one class C_E such that $f(y)=0$ in $Y-E$ and $E \in \mathfrak{B}$; then C_E is equivalent to $L_\mu^p(E)$, and, if we set $U_E(f)=U(f)$ for $f \in C_E$, U_E is linear on $L_\mu^p(E)$ and therefore can be expressed as before where the E_i of Theorem 4 will be those E_γ which, except for sets of measure zero, lie in E. But $f=0$ on $Y-E$; so we can write $U(f)=\sum_\gamma u_\gamma \bar{f}_\gamma$ if it is not identically zero, with the convention that the sum of any number of terms in which $\bar{f}_\alpha$ or u_α is zero shall be zero. Since this can be done for each $f \in L_\mu^p(Y)$, we get our final result, including the previous theorems as special cases.

THEOREM 5. *A functional U on $L_\mu^p(Y)$ is linear if and only if* (a) *$\mathfrak{A}$ is empty and U is identically zero or* (b) *$\mathfrak{A}$ is not empty and U can be expressed in the form $U(f)=\sum_\gamma u_\gamma \bar{f}_\gamma$ and $\|U\|=\sup_\gamma |u_\gamma| b_\gamma^{-1/p}$.*

If we let L^{p*} be the space of linear functionals on $L_\mu^p(Y)=L^p$, it is clear that with the given norm L^{p*} is a Banach space. To make the study of its structure simpler we introduce the class Γ of ordinal numbers as follows; If $\mathfrak{A}$ is empty, Γ is also empty; if $\mathfrak{A}$ is not empty, Γ is the class of ordinals used in ordering the sets E_γ of $\mathfrak{A}$ defined above. We use $\mathbf{x}=\{x_\gamma\}$ to stand for a real-valued function on Γ and define the special function $\mathbf{x}_\gamma$ to have the value 1 at γ and 0 elsewhere. Let $M=M(\Gamma)$ be the class of bounded functions $\mathbf{x}$ with $\|\mathbf{x}\|_M=\sup_{\gamma\in\Gamma}|x_\gamma|$; for any $q>0$ let $l^q=l^q(\Gamma)$ be the class of all $\mathbf{x}$ such that $\|\mathbf{x}\|_q=(\sum_{\gamma\in\Gamma}|x_\gamma|^q)^{1/q}<\infty$.

It is clear that the space $l^q(\Gamma)$ here defined is also the space $L_{\mu_0}^q(\Gamma)$ where μ_0 is the trivial measure assigning the measure 1 to each set containing just one point γ and $\mathfrak{X}$ is the smallest additive family on the countable sets in Γ. From Theorem 5 we have this result:

THEOREM 6. *If Γ is not empty, a functional U on l^p is linear if and only if it can be expressed as $U(\mathbf{x})=\sum_{\gamma\in\Gamma}u_\gamma x_\gamma$ where $\|U\|=\sup_{\gamma\in\Gamma}|u_\gamma|$.*

Considering again the space L^{p*}, it is clear that when $\mathfrak{A}$ is not empty the transformation which associates the element $U \in L^{p*}$ defined by $U(f)=\sum_{\gamma\in\Gamma}u_\gamma \bar{f}_\gamma$ with the element $u=\{u_\gamma b_\gamma^{-1/p}\} \in M$ is one-to-one, linear both ways and norm-preserving between all of L^{p*} and all of M; but this is precisely the definition of equivalence of two Banach spaces (Banach, p. 180); hence:

(6) L^{p*} and M are equivalent Banach spaces.

This holds even for the case in which $\mathfrak{A}$ is empty if we define $m(\Gamma)$ for Γ empty to be the Banach space with just one element, the zero element.

Slightly modifying our previous usage, we let $f_{(\gamma)}$ be $b_\gamma^{-1/p}$ on E_γ and be zero elsewhere. If we extend the definition of equivalence of Banach spaces in the obvious way to these L^p spaces we get this result:

THEOREM 7. *$L_\mu^p(\sum_{\gamma\in\Gamma}E_\gamma)$ is equivalent to $l^p(\Gamma)$ under the linear transformation for which $Tf=\mathbf{x}$, where $\mathbf{x}=\{\mathbf{x}_\gamma\}=\{\bar{f}_\gamma b_\gamma^{1/p}\}$; that is, which takes $f_{(\gamma)}$ to $\mathbf{x}_\gamma$.*

PROOF. By linearity if $f=\sum_\gamma x_\gamma f_{(\gamma)}$, $Tf=\mathbf{x}$ must be $\sum_\gamma x_\gamma \mathbf{x}_\gamma$. Now $\|f\|_p=(\sum|x_\gamma|^p)^{1/p}$ and $\|\mathbf{x}\|_p=(\sum|x_\gamma|^p)^{1/p}$; so the transformation is one-to-one on all L^p and l^p and $\|Tf\|=\|f\|$ for all $f\in L^p(\sum_\gamma E_\gamma)$; so the spaces are equivalent.

COROLLARY. *If Γ' is any set of ordinals and $\Gamma'\subset\Gamma$, then $l^p(\Gamma')$ is equivalent to the subspace $L_\mu^p(\sum_{\gamma\in\Gamma'}E)$ of $L_\mu^p(Y)$.*

Since, as is well known, $l^{1*}=m$, the following relation between L^p and l^1 is not entirely unexpected.

(7) If we define on the subspace $L_\mu^p(\sum_{\gamma\in\Gamma}E_\gamma)$ of $L_\mu^p(Y)$ the transformation T by associating to $f=\sum_{\gamma\in\Gamma}x_\gamma f_{(\gamma)}$ the function $Tf=\mathbf{x}=\sum_\gamma x_\gamma\mathbf{x}_\gamma$, then T is linear on $L_\mu^p(\sum_{\gamma\in\Gamma}E_\gamma)$ to l^1 and $\|T\|=1$.

In general, if Γ is not a finite set, the set of all Tf with $f\in L_\mu^p(\sum_{\gamma\in T}E_\gamma)$ is not all of l^1.

We conclude with some elementary remarks about the cases when Γ is a finite set; then the spaces $l^q(\Gamma)$, $q>0$, can be gotten by giving a new pseudo-norm to an n-dimensional euclidean space. When $q\geqq 1$ this pseudo-norm is actually a norm and the "spheres" $E_\epsilon[\|\mathbf{x}\|\leqq\epsilon]$ are convex point sets, and we have some of the well known Minkowski spaces. When $0<q<1$ these sets are not convex but each contains an ordinary euclidean sphere and is contained in another such sphere, so the topology in all these $l^q(\Gamma)$ is the same as that in, say, $l^2(\Gamma)$, the space with euclidean metric. Convergence in $m(\Gamma)$ is also the same as that in $l^2(\Gamma)$ when Γ is finite. However these spaces are not in general equivalent as Banach spaces; if $\Gamma=\{1\}$, all are equivalent Banach spaces; if $\Gamma=\{1,2\}$, $l^1(\Gamma)$ and $m(\Gamma)$ are equivalent; none of the others are.

INSTITUTE FOR ADVANCED STUDY

A FORMAL EXPANSION THEORY FOR FUNCTIONS OF ONE OR MORE VARIABLES[1]

NEWMAN A. HALL

It is a familiar property of the expansion of a function in series of functions that the coefficients often may be expressed in terms of the coefficients of the Taylor series for the original function. This has been done explicitly for many specific cases with functions of a single variable.[2] In this paper there is presented a method of obtaining more general results of this nature for functions of any number of variables defined by power series.

The umbral calculus introduced by Blissard in his *Theory of generic functions*[3] has been used by Lucas and Bell among others as a convenient instrument in the manipulation of generating functions. The algebra of the umbrae has been discussed by Bell[4] and some of the simplest properties of these will be used in the theory presented below.

A function $f(x)$ defined by a power series[5]

$$f(x) = \sum_n a_n \frac{x^n}{n!}$$

may be equally well defined by the matrix

$$a = \left| a_0, a_1, \cdots, a_n, \cdots \right|.$$

The umbral calculus admits the equality $a^n = a_n$, that is, the nth power of the matrix is equal to the nth term. From this it follows that $f(x) = e^{ax}$.

Functions of several variables suggest a similar notation. The function

---

[1] Presented to the Society, November 27, 1937, under the title *A formal expansion theory for functions defined by two variable power series.*

[2] N. Nielsen, *Fonctions Métasphériques*, chap. 4. N. Nielsen, *Recherches sur le développement d'une fonction analytique en series de fonctions hypergéométriques*, Annales Scientifiques d'École Normale Supérieure, (3), vol. 30 (1913), p. 12. S. Pincherle, *Alcuni teoremi sopra gli sviluppi en serie per funzioni analitiche*, Rendiconti dell' Istituto Lombardo di Scienze e Lettere, (2), vol. 15 (1882), p. 224. J. M. Whittaker, *Interpolatory Function Theory*, Cambridge, 1937.

[3] John Blissard, Quarterly Journal of Mathematics, vols. 4–6 (1861–1864).

[4] E. T. Bell, *Algebraic Arithmetic*, American Mathematical Society Colloquium Pubications, vol. 7, New York, 1927, pp. 146–159.

[5] All summations are to extend from 0 to ∞. In place of a repeated summation, $\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\cdots\sum_{n_k=0}^{\infty}A_{n_1,n_2,\cdots,n_k}$ we shall write $\sum_{n_1,\cdots,n_k}A_{n_1,\cdots,n_k}$.

$$f(x) = f(x_1, \cdots, x_k) = \sum_{n_1, \cdots, n_k} a_{n_1, \cdots, n_k} \frac{x_1^{n_1} \cdots x_k^{n_k}}{n_1! \cdots n_k!}$$

is defined also by the k-dimensional matrix

$$a = a'a'' \cdots a^{(k)} = \left| a_{i_1, \cdots, i_k} \right|.$$

The exponential property of the one-dimensional matrices may be extended by setting $a'^{n_1}a''^{n_2} \cdots a^{(k)\,n_k} = a_{n_1, n_2, \cdots, n_k}$ so that

$$f(x) = \exp \left\{ a'x_1 + a''x_2 + \cdots + a^{(k)}x_k \right\}.$$

This use of an umbral notation puts the primary emphasis on the number of the term in the series and on the coefficient rather than on the variable.

We shall call two sets of functions

$$P_n(x) = P_{n_1, \cdots, n_k}(x_1, \cdots, x_k), \qquad Q_n(y) = Q_{n_1, \cdots, n_k}(y_1, \cdots, y_k)$$

associate, if they satisfy the relation

$$\exp \left\{ x_1y_1 + x_2y_2 + \cdots + x_ky_k \right\} = \sum_{n_1, \cdots, n_k} \frac{P_n(x)Q_n(y)}{n_1! \cdots n_k!}.$$

In this discussion $P_n(x)$ and $Q_n(y)$ shall be restricted to functions defined by power series or by the matrices

$$p_n = \left| p_{n_1, \cdots, n_k; i_1, \cdots, i_k} \right|, \qquad q_n = \left| q_{n_1, \cdots, n_k; i_1, \cdots, i_k} \right|.$$

According to the theorem below, in order to expand $f(x)$ in a series of functions $P_n(x)$, it is only necessary to know the associate function $Q_n(y)$.

THEOREM I. *If $P_n(x)$ and $Q_n(y)$ are associate functions and $f(x)$ is a function defined by the matrix $a'a'' \cdots a^{(k)}$, the expansion of $f(x)$ in the series of functions $P_n(x)$ is*

$$f(x) = \exp \left\{ a'x_1 + \cdots + a^{(k)}x_k \right\} = \sum_{n_1, \cdots, n_k} \frac{P_{n_1, \cdots, n_k}(x_1, \cdots, x_k) Q_{n_1, \cdots, n_k}(a', \cdots, a^{(k)})}{n_1! \cdots n_k!}.$$

A discussion for functions of two variables will illustrate the argument required in general. From the definition of the associate functions,

$$e^{x_1y_1 + x_2y_2} = \sum_{m_1, m_2} \frac{P_{m_1, m_2}(x_1, x_2) Q_{m_1, m_2}(y_1, y_2)}{m_1! m_2!}$$

or

$$\sum_{n_1,n_2} \frac{x_1^{n_1} y_1^{n_1} x_2^{n_2} y_2^{n_2}}{n_1!n_2!} = \sum_{m_1,m_2} \frac{1}{m_1!m_2!} \sum_{r_1,r_2} \frac{p_{m_1,m_2;r_1,r_2}}{r_1!r_2!} x_1^{r_1} x_2^{r_2} \cdot \sum_{s_1,s_2} \frac{q_{m_1,m_2;s_1,s_2}}{s_1!s_2!} y_1^{s_1} y_2^{s_2}.$$

If $y_1^h y_2^k$ is replaced throughout by $a_{h,k}$, the equality exhibited is not affected in any way, since the variables in effect serve only to number the terms in the series. The equality follows from relations between the coefficients and is not dependent upon the variables. With this substitution made the theorem follows.

When one set of associate functions is known, any number of others may be derived by the use of the following result.

THEOREM II. *If $P_n(x)$ and $Q_n(y)$ are a set of associate functions and $c'c'' \cdots c^{(k)} = c_{n_1,\cdots,n_k}$ is an arbitrary matrix, then*

$$P_n(x_1c', x_2c'', \cdots, x_kc^{(k)}), \qquad Q_n(y_1/c', y_2/c'', \cdots, y_k/c^{(k)})$$

are associate functions.

Again the proof for two variables follows from an examination of the defining identity as given in the discussion of Theorem I. It is only necessary to replace $x_1^{r_1}x_2^{r_2}$ by $x_1^{r_1}x_2^{r_2}c_{r_1,r_2}$ and $y_1^{n_1}y_2^{n_2}$ by $y_1^{n_1}y_2^{n_2}/c_{n_1,n_2}$, which by reasoning as above will not affect the equality nor will it affect the left side since c_{n_1,n_2} will cancel $1/c_{n_1,n_2}$.

The associate functions may be obtained directly from the relation between the coefficients which follow from the identity defining the associate functions or more conveniently by applying Theorem II to certain basic sets. In the illustrations below it will be shown that a great number of expansions in functions of one or two variables will follow in this latter manner from two basic sets.

The Hermite polynomials suggest the more general class of functions $H_n^p(x)$, defined by the generating function

$$e^{xz - z^p/p} = \sum_n \frac{z^n}{n!} H_n^p(x),$$

so that

$$H_n^p(x) = \sum_s (-)^{s(p+1)} \frac{(-n, ps)}{(1, s)} \frac{x^{n-ps}}{p^s}.$$

The associate function $N_n^p(y)$ is readily determined from the generating function, since

$$e^{xy} = e^{xy - y^p/p + y^p/p} = \sum_n \frac{H_n^p(x) y^n e^{y^p/p}}{n!},$$

so that

$$N_n^p(y) = y^n e^{y^p/p} = \sum_r \frac{y^{n+pr}}{p^r(1, r)}.$$

The well known Neumann expansions in Bessel functions are special cases of the expansions obtained from the associate functions

$$B_n^k(\alpha, x) = \sum_r \frac{x^{n+kr}}{(\alpha + 2n/k + 1, r)r!},$$

$$L_n^k(\alpha, y) = \sum_s (-)^{ks} \frac{(-n, ks)y^{n-ks}}{(-\alpha - 2n/k + 1, s)s!}.$$

We show that these are associate by referring to the basic definition

$$\begin{aligned}
&\sum_n \frac{B_n^k(\alpha, x)L_n^k(\alpha, y)}{n!} \\
&= \sum_{n,r,s} (-)^{ks} \frac{(-n, ks)x^{n+kr}y^{n-ks}}{(\alpha + 2n/k + 1, r)(-\alpha - 2n/k + 1, s)n!r!s!} \\
&= \sum_p \frac{x^p y^p}{p!} \\
&\quad \cdot \sum_{r,s} (-)^{k(r+s)} \frac{(-p, kr)(-p + kr, ks)y^{-kr-ks}}{(-\alpha - 2p/k + 2r + 1, s)(\alpha + 2p/k - 2r + 1, r)r!s!} \\
&= \sum_{p,q} \frac{x^p y^{p-kq}}{p!q!} \\
&\quad \cdot \sum_r (-)^{kr} \frac{(-q, r)(-p, kr)(-p + kr, kq - kr)}{(-\alpha - 2p/k + 2r + 1, q - r)(\alpha + 2p/k - 2r + 1, r)r!} \\
&= \sum_{p,q} (-)^{kq} \frac{(-p, kq)x^p y^{p-kq}}{(-\alpha - 2p/k + 1, q)p!q!} \\
&\quad \cdot \sum_r \frac{(-q, r)(-\alpha - 2p/k + 1, 2r)(-\alpha - 2p/k, r)}{(-\alpha - 2p/k + q + 1, r)(-\alpha - 2p/k, 2r)r!} \\
&= \sum_{p,q} (-)^{kq} \frac{(-p, kq)x^p y^{p-kq}}{(-\alpha - 2p/k + 1, q)p!q!} \cdot \frac{(-\alpha - 2p/k + 1, q)(0, q)}{[(\frac{1}{2}\{-\alpha - 2p/k + 1\}, q)]^2} \\
&= \sum_p \frac{x^p y^p}{p!} = e^{xy}.{}^6
\end{aligned}$$

Thus B_n and L_n fulfill the definition of associate functions.

---

[6] W. N. Bailey, *Generalized Hypergeometric Series*, Cambridge, 1935, p. 25.

The Neumann series of the first kind gives expansions in

$$(\tfrac{1}{2}x)^{-v}J_{v+n}(x) = \sum_r (-)^r \frac{(\tfrac{1}{2}x)^{n+2r}}{\Gamma(v+n+r+1)r!}$$
$$= \frac{1}{\Gamma(v+n+1)} \sum_r (-)^r \frac{(\tfrac{1}{2}x)^{n+2r}}{(v+n+1, r)r!}$$
$$= \frac{(-i)^n}{\Gamma(v+n+1)} B_n^2(v; \tfrac{1}{2}ix).$$

The associate function is then

$$(i)^n\Gamma(v+n+1)L_n^2(v, -2ix)$$
$$= i^n\Gamma(v+n+1)\sum_s \frac{(-n, 2s)(-2i)^{n-2s}y^{n-2s}}{(-v-n+1, s)(1, s)}$$
$$= \Gamma(v+n+1)\sum_s (-)^s \frac{(-n, 2s)2^{n-2s}}{(-v-n+1, s)s!} y^{n-2s}$$

so that

$$f(x) = \sum_n a_n \frac{x^n}{n!}$$
$$= \sum_n (\tfrac{1}{2}x)^{-v}J_{v+n}(x) \cdot \frac{\Gamma(v+n+1)}{n!} \sum_s (-)^s \frac{(-n, 2s)2^{n-2s}}{(-v-n+1, s)s!} a_{n-2s}.$$

The Jacobi polynomial may be similarly considered. The general form is

$${}_2F_1(-n, \alpha+n; \gamma; x) = \sum_r \frac{(-n, r)(\alpha+n, r)}{(\gamma, r)(1, r)} x^r$$
$$= \sum_r (-)^n \frac{(-n, r)(\alpha+n, n-r)}{(\gamma, n-r)(1, r)} x^{n-r}$$
$$= (-)^n(\alpha+n, n)$$
$$\cdot \sum_r (-)^r \frac{(-n, r)x^{n-r}}{(-\alpha-2n+1, r)(\gamma, n-r)r!}$$
$$= (-)^n(\alpha+n, n)L_n^1(\alpha, cx), \quad c_h = 1/(\gamma, h),$$

so that according to Theorem II the associate function is

$$\frac{(-)^n}{(\alpha+n, n)} B_n^1(\alpha, y/c) = \frac{(-)^n}{(\alpha+n, n)} \sum_s \frac{(\gamma, n+s)y^{n+s}}{(\alpha+2n+1, s)s!}.$$

The Gegenbauer function, $C_n^v(x)$, includes the Legendre polynomial as the special case $v=\frac{1}{2}$. By definition, $(1-2hx+h^2)^{-v}=\sum_n h^n \cdot C_n^v(x)$. Hence

$$C_n^v(x) = \frac{(v, n)}{(1, n)} \sum_s \frac{(-n, 2s)(2x)^{n-2s}}{(-v-n+1, s)s!} = \frac{(v, n)}{(1, n)} L_n^2(v, 2x).$$

The associate function is

$$\frac{(1, n)}{(v, n)} B_n^2(v, \tfrac{1}{2}y) = \frac{(1, n)}{(v, n)} \sum_r \frac{(\frac{1}{2}y)^{n+2r}}{(v+n+1, r)r!}$$

and

$$f(x) = \sum_n a_n \frac{x^n}{n!} = \sum_n \frac{C_n^v(x)}{(v, n)} \sum_r \frac{a_{n+2r}}{2^{n+2r}(v+n+1, r)r!}.$$

The Hermite polynomial

$$H_n(x) = \sum_s \frac{(-n, 2s)}{(1, s)} \frac{x^{n-2s}}{2^s} = H_n^2(x)$$

has the associate function $N_n^2(y)=\sum_r y^{n+2r}/2^r(1, r)$, so that

$$f(x) = \sum_n a_n \frac{x^n}{n!} = \sum_n \frac{H_n(x)}{(1, n)} \sum_r \frac{a_{n+2r}}{2^r(1, r)}.$$

The Bessel functions of the third order introduced by P. Humbert[7] provide a direct extension of the Neumann expansions. By definition,

$$\begin{aligned}
&(\tfrac{1}{3}x)^{-\alpha-\beta} J_{\alpha+2n/3, \beta+n/3}(x) \\
&= \frac{1}{\Gamma(\alpha+\frac{2}{3}n+1)\Gamma(\beta+\frac{1}{3}n+1)} \\
&\quad \cdot \sum_r (-)^r \frac{(\frac{1}{3}x)^{n+3r}}{(\alpha+\frac{2}{3}n+1, r)(\beta+\frac{1}{3}n+1, r)r!} \\
&= \frac{1}{\Gamma(\alpha+\frac{2}{3}n+1)} \sum_r (-)^r \frac{(\frac{1}{3}x)^{n+3r}}{\Gamma(\beta+\frac{1}{3}(n+3r)+1)(\alpha+\frac{2}{3}n+1, r)r!} \\
&= \frac{e^{-n\pi i/3}}{\Gamma(\alpha+\frac{2}{3}n+1)} B_n^3(\alpha, cx), \qquad c_h = \frac{(\frac{1}{3}e^{\pi i/3})^h}{\Gamma(\beta+\frac{1}{3}h+1)}.
\end{aligned}$$

[7] P. Humbert, Comptes Rendus de l'Académie des Sciences, Paris, vol. 190 (1930), p. 59.

Hence the associate function is

$$
\begin{aligned}
&e^{n\pi i/3}\Gamma(\alpha+\tfrac{2}{3}n+1)L_n^{\beta}(\alpha, y/c)\\
&\quad = e^{n\pi i/3}\Gamma(\alpha+\tfrac{2}{3}n+1)\\
&\quad\quad \cdot \sum_s (-)^s \frac{(-n, 3s)y^{n-3s}}{(-\alpha-\tfrac{2}{3}n+1, s)s!}\,\frac{\Gamma(\beta+\tfrac{1}{3}(n+3s)+1)}{(\tfrac{1}{3}e^{\pi i/3})^{n-3s}}\\
&\quad = \Gamma(\alpha+\tfrac{2}{3}n+1)\Gamma(\beta+\tfrac{1}{3}n+1)\\
&\quad\quad \cdot \sum_s (-)^s \frac{(-n, 3s)y^{n-3s}}{(-\alpha-\tfrac{2}{3}n+1, s)(-\beta-\tfrac{1}{3}n+1, s)s!}\,.
\end{aligned}
$$

The simplest two-variable expansions are those in products of one-variable functions. For example, the Neumann series of Bessel functions can be extended. Thus, if

$$
f(x, y) = \sum_{m,n} a_{m,n}\,\frac{x^m}{(1, m)}\,\frac{y^n}{(1, n)},
$$

then

$$
f(x, y) = \sum_{m,n} \alpha_{m,n}\,\frac{(\tfrac{1}{2}x)^{-v}(\tfrac{1}{2}y)^{-u}}{m!n!}\,J_{v+m}(x)J_{u+n}(y),
$$

where

$$
\begin{aligned}
\alpha_{m,n} &= \Gamma(v+m+1)\Gamma(u+n+1)\\
&\quad \cdot \sum_{r,s} (-)^{r+s}\,\frac{(-m, 2r)(-n, 2s)2^{n+m-2r-2s}}{(-v-m+1, r)(-u-n+1, s)m!n!}\,a_{m-2r,n-2s}.
\end{aligned}
$$

We have here made use of the associate functions for the Bessel functions as mentioned above.

There are expansions in mixed products. Thus, for $f(x, y)$ above, we have

$$
f(x, y) = \sum_{m,n} \frac{\beta_{m,n}}{m!n!}\,H_m(x)C_n^{v}(y),
$$

where

$$
\beta_{m,n} = \frac{(1, n)}{(v, n)} \sum_{r,s} \frac{a_{m+2r,n+2s}}{2^{n+r+2s}(v+n+1, s)r!s!},
$$

giving the expansion in products of Hermite and Gegenbauer polynomials.

P. Humbert[8] introduced the set of confluent hypergeometric functions of two variables, special cases of which can be variously interpreted as extensions of the Sonine and Jacobi polynomials as well as the Bessel functions. Many of these provide expansions for two-variable functions in accordance with our theory. One of the simplest will illustrate. The polynomial

$$\phi_2(-m, -n; \gamma; x_1, x_2) = \sum_{r,s} \frac{(-m, v)(-n, s)}{(\gamma, r+s)r!s!} x_1^r x_2^s,$$

$$= (-)^{m+n} H_m^1(c'x_1) H_n^1(c''x_2),$$

$$c'^r c''^s = 1/(\gamma, r+s),$$

has the associate function

$$(-)^{m+n} N_m^1\left(\frac{y_1}{c'}, \frac{y_2}{c''}\right) = \sum_{r,s} (-)^{m+n} \frac{y_1^{m+r} y_2^{n+s}}{c'^{m+r} c''^{n+s} r!s!}$$

$$= \sum_{r,s} (-)^{m+n} \frac{(\gamma, r+s+m+n)}{(1, r)(1, s)} y_1^{m+r} y_2^{n+s}$$

so that

$$f(x_1, x_2) = \sum_{m,n} a_{m,n} \frac{x_1^m x_2^n}{m!n!}$$

$$= \sum_{m,n} (-)^{m+n} \frac{\phi_2(-m, -n; \gamma; x_1, x_2)}{(1, m)(1, n)}$$

$$\sum_{r,s} \frac{(\gamma, r+s+m+n)}{(1, r)(1, s)} a_{m+r, n+s}.$$

In addition to expansions derived from the two basic sets above there are many expansions in functions whose associate functions may be obtained by other means. As one example of this, we mention the two-variable Hermite polynomials[9] $H_{m,n}(x_1, x_2)$ defined by the expansion

$$\exp\left\{h(ax_1 + bx_2) + k(bx_1 + cx_2) - \tfrac{1}{2}(ah^2 + 2bhk + ck^2)\right\}$$

$$= \sum_{m,n} \frac{h^m k^n}{m!n!} H_{m,n}(x_1, x_2).$$

---

[8] P. Humbert, Proceedings of the Royal Society of Edinburgh, vol. 41 (1921), p. 73.

[9] Ch. Hermite, *Oeuvres*, vol. 2, pp. 293–308.

If we write $ha+kb=y_1$, $hb+kc=y_2$, then

$$e^{x_1y_1+x_2y_2} = \sum_{m,n} \frac{H_{m,n}(x_1, x_2)}{m!n!} h^m k^n \exp\left\{\tfrac{1}{2}(ah^2 + 2bhk + ck^2)\right\},$$

so that the function associate to $H_{m,n}(x_1, x_2)$ is

$$N_{m,n}(y_1, y_2) = \frac{(cy_1 - by_2)^m(-by_1 + ay_2)^n}{\Delta^{m+n}} \cdot \exp\left\{\frac{cy_1^2 + 2by_1y_2 + ay_2^2}{2\Delta}\right\}$$

where $\Delta = ac - b^2 \neq 0$.

Restrictions of space prevent our continuing with the almost endless variety of illustrations of the theory. These methods can be particularly useful in obtaining the explicit forms of new expansions. The detailed properties of the sets of functions will suffice to answer questions of convergence.

For the original suggestion of this work and continued counsel, the author wishes to express his appreciation to Harry Bateman and to E. T. Bell.

QUEENS COLLEGE

TOTAL REGULARITY OF GENERAL TRANSFORMATIONS[1]

HENRY HURWITZ, JR.

A method of summation is said to be regular if it assigns to every convergent series its actual value. If it also assigns the value $+\infty$ (or $-\infty$) to every series which diverges to $+\infty$ (or $-\infty$) it is said to be totally regular. The conditions for regularity are well known, and those for total regularity have been worked out for triangular matrix transformations by W. A. Hurwitz.[2] We here obtain necessary and sufficient conditions for total regularity for a more general type of transformation. (The former conditions, though still sufficient, are not necessary.)

Suppose $x_1, x_2, x_3, \cdots$ is the sequence of partial sums of the original series which is assumed real. A value Y is assigned to this sequence in the following way:

$$Y = \lim_{D(t)\to 0} y(t); \qquad y(t) = \sum_{k=1}^{\infty} a_k(t)\, x_k.$$

t is a variable ranging over some point set, $D(t)$ is a positive real function, and the functions $a_k(t)$ are real, but not necessarily continuous. We assume the transformation is regular so that the three Silverman-Toeplitz conditions are satisfied:

(1) $\sum_{k=1}^{\infty} |a_k(t)|$ is bounded for $D(t)$ sufficiently small;

(2) $\lim_{D(t)\to 0} \sum_{k=1}^{\infty} a_k(t) = 1$;

(3) $\lim_{D(t)\to 0} a_k(t) = 0$ for all k.

We then ask when Y will be positively infinite if $\lim_{k\to\infty} x_k = +\infty$. (We demand that for $D(t)$ sufficiently small $y(t)$ will be defined although it may be positively infinite.)

First it may be seen that for sufficiently advanced t (that is, t for which $D(t)$ is sufficiently small) there can be only a finite number of negative coefficients $a_k(t)$ in each row (that is, for each t) if the transformation is to be totally regular. Otherwise a sequence t_n with $D(t_n)\to 0$ could be picked out such that for each t_n there would be an infinite number of negative coefficients. Then a sequence x_k could be defined so that $x_k \to \infty$ and

$$\sum_{k=1}^{\infty} a_k(t_n)\, x_k$$

[1] Presented to the Society, February 24, 1940, under the title *Total regularity of infinite matrix transformations.*

[2] W. A. Hurwitz, Proceedings of the London Mathematical Society, vol. 26 (1927), p. 231.

is for all t_n, if defined at all, certainly less than, say, zero. This would be accomplished by defining x_k unusually large for some k for which $a_k(t_1)<0$, then for a k for which $a_k(t_2)<0$, and so on in the order $t_1, t_2, t_1, t_2, t_3, t_1, t_2, t_3, t_4, \cdots$, each time taking a greater k. Imposing this necessary condition considerably limits the possibilities and assures that for any $x_k\to\infty$, $y(t)$ will be defined for t sufficiently advanced.

DEFINITION. *The guard of the coefficient $a_k(t_0)$ with respect to the finite number of rows $t_1, t_2, \cdots, t_n$, written*

$$Ga_k(t_0)\big]_{t_1t_2\cdots t_n},$$

is equal to $+1$ if $a_k(t_0)\geqq 0$. If $a_k(t_0)<0$ it is equal to the maximum value of

$$\frac{a_k(t')}{|\,a_k(t_0)\,|}$$

for $t'=t_1, t_2, \cdots, t_n$.

THEOREM 1. *If a real regular transformation is to be totally regular, it is necessary and sufficient that*

(a) *for $D(t)$ sufficiently small there is only a finite number of negative coefficients in each row;*

(b) *there does not exist a sequence $t_1, t_2, \cdots$ such that $\lim_{n\to\infty} D(t_n)=0$ and for every finite number of rows $t_1, t_2, \cdots, t_F$*

$$\limsup_{n\to\infty} Ga_k^*(t_n)\big]_{t_1t_2\cdots t_F}\leqq 0$$

for some sequence $k_1^*, k_2^*, \cdots$ such that $k_n^*\to\infty$, where k_n^* may depend on F.

THEOREM 2. *Theorem* 1 *holds with* $\limsup_{n\to\infty} G$ *replaced by* $\liminf_{n\to\infty} G$.

The equivalence of these two theorems may be demonstrated directly, but it also follows by proving that the conditions of Theorem 1 are sufficient, and those of Theorem 2 are necessary for total regularity.

To prove the sufficiency part of Theorem 1, we show that if a transformation satisfies the conditions there cannot exist a sequence $x_k\to\infty$ such that there is a sequence t_n with $D(t_n)\to 0$ and $y(t_n)<M$ for all n. M is some large positive number. Suppose such sequences x_k and t_n do exist. We choose K_n such that

$$\sum_{k=1}^{K_n}|\,a_k(t_n)x_k\,|<M, \qquad \sum_{k=1}^{K_n}|\,a_k(t_n)\,|<\tfrac{1}{3}.$$

By the third S-T condition we may have $K_n \to \infty$, and indeed may assume without loss of generality that all the K_n's are so large that if k is greater than any one of them $x_k > 9M$. By the second S-T condition, we may assume that

$$\left| \sum_{k=1}^{\infty} a_k(t_n) - 1 \right| < \tfrac{1}{3}$$

for all n. By hypothesis there is some integer F such that there is no sequence $k_n^* \to \infty$ for which

$$\limsup_{n \to \infty} G a_{k_n^*}(t_n) \big]_{t_1 t_2 \cdots t_F} \leqq 0.$$

It then follows that at least for an infinite subsequence t_{n_i} every coefficient $a_k(t_{n_i})$ in the row t_{n_i} with $k > K_n$ has

$$G a_k(t_{n_i}) \big]_{t_1 t_2 \cdots t_F} > \beta$$

where β is some positive number, unique for the whole subsequence. Since $y(t_{n_1}) < M$, we must have

$$J_1 = \sum_{K_{n_1}+1}^{\infty} {}^* a_k(t_{n_1}) x_k < -M$$

where the asterisk indicates that the summation is to be extended over only those k's for which $a_k(t_{n_1}) < 0$. Because the guard with respect to the rows $t_1, t_2, \cdots, t_F$ of each of the coefficients appearing in the expression for J_1 is greater than β, we have

$$\sum_{J_1} a_k(t_{i_k}) x_k > \beta M$$

where the summation is over all the k's appearing in the expression for J_1 and the i_k's are suitably chosen integers ranging between 1 and F. We next pick from the sequence n_i an element n_2' such that $K_{n_2'}$ is greater than any k appearing in J_1. As before we have

$$J_2 = \sum_{K_{n_2'}+1}^{\infty} {}^* a_k(t_{n_2'}) x_k < -M$$

so that

$$\sum_{J_1, J_2} a_k(t_{i_k}) x_k > 2\beta M$$

where the summation is now over all the k's appearing in either J_1 or J_2. Proceeding in this way we can prove that it is possible to pick from the rows $t_1, t_2, \cdots, t_F$ a series diverging to positive infinity. But

since there is only a finite number of negative terms in each row it is then impossible that

$$y(t_i) < M$$

for $i = 1, \cdots, F$, so that an x_k sequence of the type assumed cannot exist.

To prove that the conditions of Theorem 2 are necessary, we assume that a sequence t_n of the kind forbidden by (b) of Theorem 1, with lim sup G replaced by lim inf G, exists and actually construct a sequence $x_k \to \infty$ such that

$$\lim_{D(t)\to 0} y(t) \neq +\infty.$$

By the second S-T condition we may assume

$$\sum_{k=1}^{\infty} |a_k(t_n)| < T > 1$$

for all n. Let $n_1 = 1$. Take K_1 such that

$$\sum_{k=K_1+1}^{\infty} |a_k(t_{n_1})| < \frac{1}{2 \cdot 2^2} = \frac{1}{8}.$$

Define

$$x_k = \frac{1}{T}, \qquad k = 1, 2, \cdots, K_1.$$

By the assumption we have in general, for $s > 1$,

$$\liminf_{n\to\infty} G a_k^{(s-1)}(t_n)]_{t_{n_1} t_{n_2} \cdots t_{n_{s-1}}} \leqq 0$$

for some sequence $k_n^{(s-1)}$ which becomes infinite with n. We can therefore choose $n_s > n_{s-1}$ so that

(a) $G a_{k_{n_s}}^{(s-1)}(t_{n_s})]_{t_{n_1} t_{n_2} \cdots t_{n_{s-1}}} < 1/s \cdot 2^s$,

(b) $k_{n_s}^{(s-1)} > K_{s-1}$,

(c) $D(t_{n_s})$ is so small that $\sum_{k=1}^{K_{s-1}} |a_k(t_{n_s}) x_k| < 1$.

Call $k_{n_s}^{(s-1)}$ simply k_s. Take $K_s > k_s$ such that

$$\sum_{k=K_s+1}^{\infty} |a_k(t_{n_p})| < \frac{1}{(s+1)2^{s+1}}$$

for $p = 1, 2, \cdots, s$. Define in general

$$x_k = s/T, \qquad K_{s-1} + 1 \leqq k \leqq K_s \quad (k \neq k_s), \qquad x_{k_s} = X_s$$

where X_s is the greater of $s/|a_{k_s}(t_{n_s})|$, s/T. It is apparent that the x_k sequence thus defined approaches infinity. Also, since $a_{k_s}(t_{n_s})$ is negative and has a small guard, it may be verified that $y(t_{n_s})$ remains bounded for all s, so that the transformation is not totally regular.

THEOREM 3. *Theorems* 1 *and* 2 *hold with the modification that the sequence* k_n^* *must be independent of* F.

To prove this it must be shown that if a sequence t_n of the type forbidden by Theorems 1 or 2 exists, then one with k_n^* independent of F can be found. But this can easily be done by taking a subsequence $t_{n_1}, t_{n_2}, t_{n_3}, \cdots$ from the sequence $t_1, t_2, t_3, \cdots$ so that some negative coefficient $a_{k_i'}(t_{n_i})$ with large k_i' in row t_{n_i} has a small guard with respect to all the rows $t_{n_1}, t_{n_2}, \cdots, t_{n_{i-1}}$. The k_i' and t_{n_i} sequences thus obtained will be of the type forbidden by Theorem 3. The same line of reasoning also serves to prove a second modification:

THEOREM 4. *Theorem* 3 *will hold with*

$$\limsup_{n\to\infty} Ga_{k_n^*}(t_n)\big]_{t_1 t_2 \cdots t_{n-1}} \leqq 0$$

instead of

$$\limsup_{n\to\infty} Ga_{k_n^*}(t_n)\big]_{t_1 t_2 \cdots t_F} \leqq 0.$$

(*The same thing holds with inferior limits.*)

The most powerful combination of these four theorems is obtained by using the necessary conditions of Theorem 2 and the sufficient conditions (in terms of superior limits) of Theorem 4. The theorem obtained by W. A. Hurwitz for triangular matrix transformations follows quite easily from these results.

THEOREM 5. *It is necessary (but not sufficient) for a real regular transformation to be totally regular that there does not exist a sequence of coefficients* $a_{k_n}(t_n)$, $n = 1, 2, \cdots$, *such that* (a) $a_{k_n}(t_n) < 0$ *for all* n; (b) *the maximum number of* n'*s for which* k_n *has any particular value is finite*; (c) $\lim_{n\to\infty} D(t_n) = 0$; (d) $\sum_{n=1}^{\infty} a_{k_n}(t_n) = -\infty$.

THEOREM 6. *It is necessary for a real regular transformation to be totally regular that*

$$\lim_{D(t)\to 0} \sum_{k=1}^{\infty} [\,|\,a_k(t)\,| - a_k(t)] = 0.$$

HARVARD UNIVERSITY

A NEW APPROACH TO THE CRITICAL VALUE THEORY

W. MAYER

The well known inequalities of M. Morse have as algebraical foundation the rank-equations

$$rB_i(\Sigma - \Sigma_1) = rB_i(\Sigma) - rB_i(\Sigma_1) + rD_i(\Sigma_1, \Sigma) + rD_{i-1}(\Sigma_1, \Sigma), \quad i = 0, 1, 2, \cdots . \tag{I}$$

These formulas hold for any topological group-system Σ, any subsystem Σ_1 of Σ, and the difference-system $\Sigma - \Sigma_1$. (W. Mayer, *Topologische Gruppensysteme*, Monatshefte für Mathematik und Physik, vol. 47 (1938); henceforth referred to as M, TG.) Here $B_i(\Sigma)$ denotes the i-dimensional Betti group of Σ, while $B_i(\Sigma_1)$ and $B_i(\Sigma - \Sigma_1)$ are these groups for Σ_1 and $\Sigma - \Sigma_1$ respectively. The symbol $r(\ )$, of course, stands for the rank of the group in the parentheses. By $D_i(\Sigma_1, \Sigma)$ we mean the subgroup of $B_i(\Sigma_1)$ containing all the classes of this group whose elements bound in Σ.

The formula (I) was first derived for the case of a complex in Lefschetz' *Topology*, 1930 (p. 150), and independently for the complex modulo 2 by J. Rybarz, Monatshefte für Mathematik und Physik (1931).

In the generality needed here the proof of (I) is given in M, TG (pp. 54–57), under the assumption, of course, that all the ranks appearing in (I) are finite, since otherwise the formula would be meaningless. But the proof there given shows also that

(a) $rB_i(\Sigma - \Sigma_1) = \infty$ implies that either $rB_i(\Sigma)$ or $rD_{i-1}(\Sigma_1, \Sigma)$, or both, are infinite;

(b) $rB_i(\Sigma - \Sigma_1)$ finite implies $rD_{i-1}(\Sigma_1, \Sigma)$ finite, and if in addition $rB_i(\Sigma_1)$ is finite then $rB_i(\Sigma)$ is finite too; and

(c) $rB_i(\Sigma - \Sigma_1) = 0$ implies $rD_{i-1}(\Sigma_1, \Sigma) = 0$ and if in addition $rB_i(\Sigma_1)$ is finite, then $rB_i(\Sigma_1) = rB_i(\Sigma) + rD_i(\Sigma_1, \Sigma)$.

As an immediate consequence of equations (I) we notice the inequality

$$rB_i(\Sigma) \leqq rB_i(\Sigma_1) + rB_i(\Sigma - \Sigma_1), \tag{I'}$$

which is true whenever the terms on the right are finite (remark (b)) and trivial otherwise. The next step in attaining the Morse inequalities is the application of (I) to $m+2$ topological group-systems satisfying the inclusion relations

$$\Sigma_m \supset \Sigma_{m-1} \supset \cdots \supset \Sigma_0 \supset \Sigma_{-1} \tag{1}$$

with Σ_{-1} empty (all $L_i(\Sigma_{-1})$ being zero-groups). Assuming that all ranks involved are finite, we get by summation on k in

$$rB_i(\Sigma_k - \Sigma_{k-1}) = rB_i(\Sigma_k) - rB_i(\Sigma_{k-1}) + rD_i(\Sigma_{k-1}, \Sigma_k) + rD_{i-1}(\Sigma_{k-1}, \Sigma_i) \tag{2}$$

the formula

$$\sum_{k=0}^{m} rB_i(\Sigma_k - \Sigma_{k-1}) = rB_i(\Sigma_m) + \sum_{k=1}^{m} rD_i(\Sigma_{k-1}, \Sigma_k) + rD_{i-1}(\Sigma_{k-1}, \Sigma_k), \tag{3}$$

since all the $rB_i(\Sigma_{-1})$, $rD_i(\Sigma_{-1}, \Sigma_0)$ are zero, Σ_{-1} being empty.

Let Σ_m be a neighborhood-space and let Σ_k, $k = m-1, \cdots, -1$, be subspaces of Σ_m satisfying the inclusion relation (1). With the introduction of singular simplices and chains modulo 2 each space Σ_k gives rise to a topological group system, which we also denote by Σ_k. (The i-dimensional complexes K^i are the finite chains of singular simplices.) The relations (1) are then inclusion relations for the so-constructed group-systems Σ_k, and for these systems relation (3) will hold provided the ranks involved are finite. The finiteness of all these ranks will follow from the finiteness of those of the left side of (3), that is, of the

$$rB_i(\Sigma_k - \Sigma_{k-1}).$$

If these are finite, we see from (2) that $rB_i(\Sigma_0)$ $(= rB_i(\Sigma_0 - \Sigma_1))$ is finite, and thus from (2) and remark (b) we conclude that $rB_i(\Sigma_1)$ is finite. So, step-by-step, using (2) and remark (b) we find that all the ranks $rB_i(\Sigma_k)$ are finite. Since $D_i(\Sigma_{k-1}, \Sigma_k) \subset B_i(\Sigma_{k-1})$, the ranks of the groups D_i are finite too. We have then the result:

The equations (3) *hold for the group-systems* Σ_k *if only the ranks appearing on the left sides of these equations are finite.*

It is an interesting fact, in view of its geometrical implications, that the finiteness of the ranks $rB_i(\Sigma_k - \Sigma_{k-1})$ has as a consequence the finiteness of the Betti numbers $rB_i(\Sigma_k)$. As an additional remark it may be noticed that equations (3) can be written in the form of Morse's inequalities (of the strong type) if (3) holds for all indices i. Denoting by M_i the left sides of (3):

$$M_i = \sum_{k=0}^{m} rB_i(\Sigma_k - \Sigma_{k-1}), \tag{4}$$

and successively eliminating the sums with terms rD_i, we get

$$
\begin{aligned}
M_0 &\geqq R_0, \\
M_1 - M_0 &\geqq R_1 - R_0, \\
M_2 - M_1 + M_0 &\geqq R_2 - R_1 + R_0, \\
&\cdots\cdots\cdots\cdots\cdots\cdots, \\
M_n - M_{n-1} + \cdots \pm M_0 &= R_n - R_{n-1} + \cdots \pm R_0,
\end{aligned}
\tag{3'}
$$

where n is the dimension of the topological system Σ_m and R_i stands for $rB_i(\Sigma_m)$, the ith Betti number of Σ_m. The inequalities (3'), however, are not fully equivalent to equations (3) since they result only if (3) holds for all the dimensions i from zero up to n, the dimension of Σ_m. In addition, in (3) the difference

$$(M_i - M_{i-1} + \cdots \pm M_0) - (R_i - R_{i-1} + \cdots \pm R_0)$$

is shown to be $\sum rD_i(\Sigma_{k-1}, \Sigma_k)$, which fact, of course, is of importance since it readily yields the conditions for the equality sign in (3').

A subdivision of the space Σ_m by subspaces Σ_k of Σ_m can be obtained by the use of a real bounded function $\mathfrak{F}$ defined on Σ_m. Let σ_0 and σ_m be the greatest lower and least upper bound of $\mathfrak{F}$ respectively, so that

$$\sigma_m \geqq \mathfrak{F} \geqq \sigma_0; \tag{5}$$

then intermediate values σ_k, $k=1, \cdots, m-1$, and $\sigma_{-1}<\sigma_0$ may be introduced and subjected to the inequalities

$$\sigma_m > \sigma_{m-1} > \cdots > \sigma_1 > \sigma_0 > \sigma_{-1} \tag{6}$$

if $\mathfrak{F}$ is not constant on Σ_m. If $\mathfrak{F}$ is constant, (6) can be reduced to $\sigma_0>\sigma_{-1}$.

We now define the subspace Σ_k as the set of all points P of Σ_m such that $\mathfrak{F}(P) \leqq \sigma_k$, that is, symbolically,

$$\Sigma_k \equiv \{\mathfrak{F} \leqq \sigma_k\}. \tag{7}$$

(We observe that in this notation the subspace Σ_m defined in (7) coincides with the space originally so denoted.)

Constructing the corresponding topological group-systems, denoted likewise by Σ_m (with Σ_{-1} consisting of zero groups only), we see that equations (3) hold if only the sum M_i of (4) is finite. This shows the full generality of these equations, in which only the concepts of neighborhood space and its partition by subspaces (not necessarily defined by a function $\mathfrak{F}$) are used. (Later on the validity of formula

(3) will be shown for a partition of the space by a countable set of subspaces.)

Let Σ_m be a compact Riemannian manifold and $\mathfrak{F}$ a function of class C^2 defined on Σ_m and having a finite number only of stationary points. Corresponding to these points there are a finite number of stationary values σ_k, $k=0, 1, \cdots, m$, with σ_0 and σ_m the absolute minimum and maximum respectively. Using these values in defining the Σ_k as previously described and denoting by the same letters Σ_k the corresponding group-systems, we again arrive at equations (3), given the finiteness of their left sides M_i.

Denote by Σ_k' the space of all points P of Σ_m with the property $\mathfrak{F}(P)<\sigma_k$; then obviously $\Sigma_k' \subset \Sigma_k$.

By definition (Seifert-Threlfall, *Variationsrechnung im Grossen*, §4; we refer to this henceforth as S.T.) the type numbers $m_i(\sigma)$ for any value σ of $\mathfrak{F}$ are the ranks of the Betti groups of $\Sigma_{(\sigma)}-\Sigma_{(\sigma)}'$, where $\Sigma_{(\sigma)}$ and $\Sigma_{(\sigma)}'$ are the point sets $\{\mathfrak{F}\leqq\sigma\}$ and $\{\mathfrak{F}<\sigma\}$ respectively. The value σ is called critical if some of the $m_i(\sigma)$ are different from zero.

Only a stationary value (that is, a value belonging to a stationary point) can be critical, so that only for the stationary values σ_k, $k=0, 1, \cdots, m$, are the type numbers

$$m_i(\sigma_k) = rB_i(\Sigma_k - \Sigma_k') \tag{8}$$

not all necessarily zero. In consequence of our assumption all the stationary points are *isolated* and thus their contribution to the corresponding $m_i(\sigma_k)$ will be finite (S.T., §10). Since only a finite number of stationary points belong to a stationary value σ_k, $m_i(\sigma_k)$ *will be finite* (S.T., p. 87).

If therefore we can prove that

$$rB_i(\Sigma_k - \Sigma_k') = rB_i(\Sigma_k - \Sigma_{k-1}), \tag{9}$$

then not only is (3) verified, but its left side is shown to be the sum of all the type numbers of dimension i. Thus, in proving isomorphisms between the Betti groups of the systems

$$(\alpha_k) \qquad \Sigma_k - \Sigma_k'$$

and

$$(\beta_k) \qquad \Sigma_k - \Sigma_{k-1}$$

for $k=0, 1, \cdots, m$, we likewise prove (9) and thereby (3).

The above isomorphisms are included in the following two statements:

(a) Each class of $B_i(\Sigma_k-\Sigma_{k-1})$ lies in a definite class of $B_i(\Sigma_k-\Sigma_k')$.

(b) Each class of $B_i(\Sigma_k - \Sigma_k')$ contains one and only one class of $B_i(\Sigma_k - \Sigma_{k-1})$.

Indeed, (a) and (b) establish a correspondence between the classes of the above groups which is obviously an isomorphism.

As to the proof of the statements, (a) is almost self-evident, since an i-cycle of $\Sigma_k - \Sigma_{k-1}$ is an i-cycle of $\Sigma_k - \Sigma_k'$, and an i-cycle of $\Sigma_k - \Sigma_{k-1}$ which is homologous to zero in $\Sigma_k - \Sigma_{k-1}$ is easily seen to be an i-cycle of $\Sigma_k - \Sigma_k'$ homologous to zero in $\Sigma_k - \Sigma_k'$.

We now prove the first part of (b), that is, each class of $B_i(\Sigma_k - \Sigma_k')$ contains a class of $B_i(\Sigma_k - \Sigma_{k-1})$. This is by no means evident, and to prove it we must make use of $\mathfrak{F}$-deformations, which in the case we are considering will exist. Suppose given $\{Z^i\} \subset B_i(\Sigma_k - \Sigma_k')$; then there exists a $K^{i-1} \subset \Sigma_k'$ such that

$$R(Z^i) = K^{i-1}. \tag{10}$$

But, on the compact point set K^{i-1}, the continuous function $\mathfrak{F}$ has somewhere a maximum which, of course, is smaller than σ_k, $(K^{i-1} \subset \Sigma_k')$. Thus K^{i-1} lies in some Σ^* defined by $\{\mathfrak{F} \leqq \sigma_k - \epsilon\}$, where $\epsilon > 0$ may be so chosen that $\sigma_k - \epsilon > \sigma_{k-1}$. Then (S.T., p. 87) an $\mathfrak{F}$-deformation exists which carries the point set $\{\mathfrak{F} \leqq \sigma_k - \epsilon\}$ into the point set $\{\mathfrak{F} \leqq \sigma_{k-1}\}$. Thus there exist a K^i on $\{\mathfrak{F} \leqq \sigma_k - \epsilon\}$ and a K_1^{i-1} on $\{\mathfrak{F} \leqq \sigma_{k-1}\}$ such that

$$R(K^i) = K^{i-1} + K_1^{i-1}. \tag{11}$$

Adding (10) to (11) we get

$$R(Z^i + K^i) = K_1^{i-1}, \tag{12}$$

thus showing that $Z^i + K^i$, a cycle of the class $\{Z^i\}$ of $B_i(\Sigma_k - \Sigma_k')$, is a cycle of $\Sigma_k - \Sigma_{k-1}$. Hence $\{Z^i\}$ contains this cycle of $\Sigma_k - \Sigma_{k-1}$, and, according to (a), the class of $B_i(\Sigma_k - \Sigma_{k-1})$ it represents.

We finally prove the second part of (b), namely: Each class of $B_i(\Sigma_k - \Sigma_k')$ contains only one class of $B_i(\Sigma_k - \Sigma_{k-1})$. This again follows from the statement: If a cycle Z^i of $\Sigma_k - \Sigma_{k-1}$ considered as a cycle of $\Sigma_k - \Sigma_k'$ is homologous to zero in $\Sigma_k - \Sigma_k'$, then this cycle will be homologous to zero in $\Sigma_k - \Sigma_{k-1}$. For this proof one must again make use of the $\mathfrak{F}$-deformation mentioned above. Let Z^i be a cycle of $\Sigma_k - \Sigma_{k-1}$ homologous to zero in $\Sigma_k - \Sigma_k'$; then there exist a K^{i+1} in Σ_k and a K^i in Σ_k' such that

$$R(K^{i+1}) = Z^i + K^i, \tag{13}$$

which shows that $Z^i \sim 0$ in $\Sigma_k - \Sigma_k'$. Furthermore there will exist a K^{i-1} in Σ_{k-1} such that

$$R(Z^i) = K^{i-1}, \tag{14}$$

since Z^i is a cycle of $\Sigma_k - \Sigma_{k-1}$. From (13) and (14) follows

$$R(K^i) = K^{i-1}. \tag{15}$$

The compact set K^i of Σ_k, $(\mathfrak{J} < \sigma_k)$, will lie in some $\{\mathfrak{J} \leqq \sigma_k - \epsilon\}$. Using the $\mathfrak{J}$-deformation carrying $\{\mathfrak{J} \leqq \sigma_k - \epsilon\}$ into $\{\mathfrak{J} \leqq \sigma_{k-1}\}$, we establish the existence of a chain K_1^{i+1} of $\{\mathfrak{J} \leqq \sigma_k - \epsilon\}$ and two chains K_1^i and L_1^i of $\{\mathfrak{J} \leqq \sigma_{k-1}\}$ satisfying the relation

$$R(K_1^{i+1}) = K^i + K_1^i + L_1^i, \tag{16}$$

where L_1^i, as the deformation chain of $R(K^i) = K^{i-1}$ (which lies in Σ_{k-1}) will lie in Σ_{k-1}, by definition of an $\mathfrak{J}$-deformation. Adding the two relations (13) and (16), we get finally

$$R(K^{i+1} + K_1^{i+1}) = Z^i + K_1^i + L_1^i, \tag{17}$$

where $K_1^i + L_1^i \subset \Sigma_{k-1}$ and $K^{i+1} + K_1^{i+1} \subset \Sigma_k$. This proves that $Z^i \sim 0$ in $\Sigma_k - \Sigma_{k-1}$.

We add that in proving (9) we made no use of the fact that either of the values σ_k or σ_{k-1} was stationary. We used only the existence of the above mentioned $\mathfrak{J}$-deformations, and these will always exist provided only that between σ_k and σ_{k-1} there are no stationary values. Given this condition, (9) always holds (of course, in the form $0 = 0$ when σ_k is not a stationary value).

The theory so far developed is unable to deal with the concept of the "variational calculus in the large," since in this case the function $\mathfrak{J}$ defined on Σ is unbounded above. We now adapt the theory to that case.

First we prove the following lemma:

LEMMA. *For three group-systems* Σ_α, $\alpha = 1, 2, 3$, *satisfying the inclusion relations*

$$\Sigma_1 \subset \Sigma_2 \subset \Sigma_3, \tag{18}$$

the relation

$$rD_i(\Sigma_1, \Sigma_3) \leqq rD_i(\Sigma_1, \Sigma_2) + rD_i(\Sigma_2, \Sigma_3) \tag{19}$$

holds.

PROOF. Each class of $D_i(\Sigma_1, \Sigma_3)$ is contained in a definite class of $D_i(\Sigma_2, \Sigma_3)$ since each cycle of Σ_1 bounding in Σ_3 is a cycle of Σ_2 bounding in Σ_3 and each cycle of Σ_1 bounding in Σ_1 is a cycle of Σ_2 bounding in Σ_2. By correlating with an element of $D_i(\Sigma_1, \Sigma_3)$ that

element of $D_i(\Sigma_2, \Sigma_3)$ which contains it, we define a homomorphism,

$$D_i(\Sigma_1, \Sigma_3) \to D_i(\Sigma_2, \Sigma_3), \tag{20}$$

whose kernel group is the subgroup of $D_i(\Sigma_1, \Sigma_3)$ consisting of all classes bounding in Σ_2, hence $D_i(\Sigma_1, \Sigma_2)$. Denoting the map-group of the homomorphism by $D_i'(\Sigma_2, \Sigma_3)$ we have the isomorphism

$$D_i(\Sigma_1, \Sigma_3) - D_i(\Sigma_1, \Sigma_2) \approx D_i'(\Sigma_2, \Sigma_3) \tag{21}$$

and the rank-equation

$$rD_i(\Sigma_1, \Sigma_3) = rD_i(\Sigma_1, \Sigma_2) + rD_i'(\Sigma_2, \Sigma_3), \tag{22}$$

whence (19) follows (since $D_i'(\Sigma_2, \Sigma_3) \subset D_i(\Sigma_2, \Sigma_3)$).

Combining (I) and (19) we derive

$$rB_i(\Sigma_3 - \Sigma_1) \leqq rB_i(\Sigma_3 - \Sigma_2) + rB_i(\Sigma_2 - \Sigma_1), \tag{23}$$

thus showing that in subdividing a given partition we never decrease the sum M_i.

As for the validity of the proof given for (23), we have to establish the legality of the use of formula (I). This use will indeed be legitimate if the rank equations (I) hold for $B_i(\Sigma_3 - \Sigma_2)$ and $B_i(\Sigma_2 - \Sigma_1)$, since then all ranks in (I) for $\Sigma = \Sigma_3$ are finite, by remark (a) of page 838 and equation (19), and the employment of (I) is thus justified for all the three Betti groups appearing in (23). This condition will in fact be satisfied whenever (23) is used in the sequel.

Equation (23), however, is true without restrictions, as we show by the use of a formula of a more general type, namely, (4′) of page 46 in M, TG. By this formula, (18) implies the isomorphism $\Sigma_3 - \Sigma_2 \approx (\Sigma_3 - \Sigma_1) - (\Sigma_2 - \Sigma_1)$ which generalizes a well known group-isomorphism to group-systems. Using (I′) for the replacement of Σ and Σ_1 by $\Sigma_3 - \Sigma_1$ and $\Sigma_2 - \Sigma_1$ respectively we again arrive at formula (23), but now without any restriction.

After this remark we extend the proof of formula (3). Let the space Σ be subdivided by a countable set of subspaces Σ_k, $k = -1, 0, 1, \cdots$, satisfying the inclusion relation

$$\Sigma_{-1} \subset \Sigma_0 \subset \Sigma_1 \subset \cdots \subset \Sigma_m \subset \cdots \subset \Sigma \tag{24}$$

with Σ_{-1} empty and $\sum_{m=0}^{\infty}(\Sigma_m) = \Sigma$ (we write $\Sigma = \Sigma_\infty$). Then formula (3) holds for an infinite but bounded sum

$$M_i = \sum_{k=1}^{\infty} rB_i(\Sigma_k - \Sigma_{k-1}) \tag{25}$$

if only the partition satisfies an additional assumption on its topological nature, namely:

(C) *Any compact point set of* Σ *lies in some* Σ_k $(k \neq \infty)$.

As a matter of fact, (C) is satisfied if the partition (24) is defined by the level surfaces of a continuous real function $\mathfrak{F}$ unbounded above, the subspace Σ_k being the set of all points P such that $\mathfrak{F}(P) \leqq \sigma_k$ and $\sigma_{-1} < \sigma_0 < \sigma_1 < \cdots$ being an unbounded sequence, with $\sigma_{-1} < \mathfrak{F}(P)$ for all points P of Σ.

Since by assumption M_i is finite, the left side of (3) will be finite, and thus (3) holds for any m. Thus each sum

$$\sum_{k=1}^{m} rD_i(\Sigma_{k-1}, \Sigma_k), \qquad \sum_{k=1}^{m} rD_{i-1}(\Sigma_{k-1}, \Sigma_k) \tag{26}$$

in (3), being not larger than M_i, converges as $m \to \infty$. As a further consequence, $rB_i(\Sigma_m)$ also converges as $m \to \infty$:

$$\lim_{m \to \infty} rB_i(\Sigma_m) = R_i'. \tag{27}$$

If only we can show that $R_i' = rB_i(\Sigma)$, we shall have proved formula (3) for $m = \infty$, that is,

$$M_i = R_i + \sum_{k=1}^{\infty} rD_i(\Sigma_{k-1}, \Sigma_k) + \sum_{k=1}^{\infty} rD_{i-1}(\Sigma_{k-1}, \Sigma_k), \tag{28}$$

writing $R_i = rB_i(\Sigma)$.

Due to the convergence of the sequences (26) and (27), all of whose terms—as ranks—are positive integers, or zeros, there exists an integer N such that for $k \geqq N$

$$\begin{aligned} rB_i(\Sigma_k - \Sigma_{k-1}) &= 0, \qquad rD_i(\Sigma_{k-1}, \Sigma_k) = 0, \\ rD_{i-1}(\Sigma_{k-1}, \Sigma_k) &= 0, \qquad rB_i(\Sigma_k) - R_i' = 0. \end{aligned} \tag{29}$$

Consequently, for $h > k > N$ the right-hand terms of the inequality

$$rB_i(\Sigma_h - \Sigma_k) \leqq \sum_{j=k+1}^{h} rB_i(\Sigma_j - \Sigma_{j-1}) \tag{30}$$

are zero, so that

$$rB_i(\Sigma_h - \Sigma_k) = 0. \tag{31}$$

From (31) and the assumption (C) we easily conclude that for $k > N$

$$rB_i(\Sigma - \Sigma_k) = 0. \tag{32}$$

In fact, (32) is proved if, for any C^i representing an element of $B_i(\Sigma - \Sigma_k)$, we can show that $C^i \sim 0$ in $\Sigma - \Sigma_k$. As a compact set on Σ, ($C^i \subset \Sigma$ with $R(C^i) \subset \Sigma_k$), C^i lies in some $\Sigma_h \supset \Sigma_k$, and thus represents an element of $B_i(\Sigma_h - \Sigma_k)$. But since this group has the rank zero, by (31), $C^i \sim 0$ in $\Sigma_h - \Sigma_k$. Therefore there exists a K^{i+1} and a K^i satisfying

$$K^{i+1} \subset \Sigma_h, \qquad K^i \subset \Sigma_k, \tag{33}$$

such that

$$R(K^{i+1}) = C^i + K^i; \tag{34}$$

this shows that $C^i \sim 0$ in $\Sigma - \Sigma_k$ and proves (32).

Using (32) in connection with the main formula (I) we have for $k > N$

$$0 = rB_i(\Sigma) - rB_i(\Sigma_k) + rD_i(\Sigma_k, \Sigma) + rD_{i-1}(\Sigma_k, \Sigma). \tag{35}$$

Remark. The use of (I) is again justified, since all ranks involved are finite. (For $rB_i(\Sigma)$ and $rD_{i-1}(\Sigma_k, \Sigma)$ see remark (b) on page 838; remember also that $rB_i(\Sigma_k) = R_i'$ and $rD_i(\Sigma_k, \Sigma) \leqq rB_i(\Sigma_k)$.) According to our remark (c) on page 838 we have $rD_{i-1}(\Sigma_k, \Sigma) = 0$ in formula (35). *We shall prove that in addition* $rD_i(\Sigma_k, \Sigma) = 0$.

This again is shown if for any C^i representing an element of $D_i(\Sigma_k, \Sigma)$ we can prove that $C^i \sim 0$ in Σ_k. Let C^i be such an element; then there exists a K^{i+1} such that

$$C^i = R(K^{i+1}), \qquad C^i \subset \Sigma_k. \tag{36}$$

But, again, the compact set K^{i+1} lies in some $\Sigma_h \supset \Sigma_k$, and thus C^i represents an element of $D_i(\Sigma_k, \Sigma_h)$, whose rank is not larger than

$$\sum_{j=k}^{h-1} rD_i(\Sigma_j, \Sigma_{j+1}) = 0.$$

Therefore $rD_i(\Sigma_k, \Sigma_h) = 0$, so that C^i is homologous to zero in Σ_k, which was to be proved.

This shown, (35) reduces to

$$rB_i(\Sigma_k) = rB_i(\Sigma), \qquad k > N, \tag{37}$$

thus proving (28).

If for the concept of variational calculus in the large ($f(x, x')$ positive definite and of class C^2) we restrict ourselves to the case where only a finite number of stationary points (extremals) are below any $\mathfrak{I}$-level ($\mathfrak{I} = \int_p^q f(x, x')dt > 0$) and if the sequence σ_k, $k = 0, 1, \cdots$, un-

bounded above, includes all the stationary values, then $\mathfrak{Z}$-deformations exist (S.T., p. 66), and can be used to prove homomorphisms between the systems (α_k) and (β_k) on page 841 in exactly the same way as before.

Hence (9) will hold, and again the type numbers

$$m_i(\sigma_k) = rB_i(\Sigma_k - \Sigma_k') \tag{38}$$

are finite (S.T., §14, Theorem II and §17, Theorem II), so that (3) is true for any m.

In taking the sum $M_i = \Sigma m_i(\sigma_k)$, only critical values σ_k (by their definition) add nonzero terms, and if noncritical values σ were used in the construction, they can be omitted in writing M_i. Hence for this case too the Morse equations hold for a finite M_i.

Remark. If, in the case just considered, for any two subspaces Σ', Σ'' of Σ the ranks of $D_i(\Sigma', \Sigma'')$ and $D_{i-1}(\Sigma', \Sigma'')$ are zero ($\Sigma' \subset \Sigma''$), then $M_i = R_i$. For finite M_i, of course, this follows from (28), but for infinite M_i we have to prove the statement. First, from (I)—or more exactly from the proof leading to (I)—we have, for any[1] k

$$rB_i(\Sigma) \geqq rB_i(\Sigma_k); \tag{39}$$

then for any m we have

$$\sum_{k=0}^{m} rB_i(\Sigma_k - \Sigma_{k-1}) = rB_i(\Sigma_m). \tag{40}$$

Since the left side of (40) diverges for $m \to \infty$, so does $rB_i(\Sigma_m)$ diverge, and hence, by (39), $rB_i(\Sigma)$ cannot be finite.

INSTITUTE FOR ADVANCED STUDY

[1] In fact, we have $B_i(\Sigma - \Sigma_k) \approx B_i(\Sigma) - F_i(\Sigma, \Sigma_k)$, $F_i(\Sigma, \Sigma_k) \approx B_i(\Sigma_1)$ (M, TG, pp. 55–57).

A THEOREM CONCERNING CLOSED AND COMPACT POINT SETS WHICH LIE IN CONNECTED DOMAINS[1]

HARLAN C. MILLER

The purpose of this paper is to show that the following theorem holds in any space which satisfies Axioms 0, 1, and 2 of R. L. Moore's *Foundations of Point Set Theory*.[2]

If g denotes a point set, $\bar{g}$ will be used to denote the set g together with all its limit points. For each positive integer n, G_n will denote the collection G_n of Axiom 1.

THEOREM. *If M is a closed and compact subset of a connected domain D, then there exists a compact continuum containing M and lying in D.*

PROOF. For each point P of D, there exists a region g_P of G_1 containing P such that $\bar{g}_P$ is a subset of D. By Axiom 2, there exists a connected domain d_P containing P which is a subset of g_P. Let U_1 denote the collection of all domains d_P for each point P of D. The point set M is closed and compact, and hence, by Theorem 22 of Chapter I, it is covered by a finite subcollection W_1 of U_1. By Theorem 77 of Chapter I, for each pair of domains x and y of W_1 there exists a simple chain xy whose links are domains of U_1 and whose first and last links are x and y respectively. Let V_1 denote the collection of all domains v such that for some two domains x and y of W_1, v is a link of the chain xy. The sum of all the domains of the finite collection V_1 is a connected domain D_1. Similarly, there exists a finite collection V_2 of connected domains such that if v is any domain of V_2, then $\bar{v}$ is a subset of some region of G_2 and of some domain of V_1, and such that the sum of the domains of V_2 is a connected domain D_2. This process can be continued. Thus there exists an infinite sequence $V_1, V_2, V_3, \cdots$ such that, for each n, (1) V_{n+1} is a finite collection of connected domains such that if v is any one of them then $\bar{v}$ is a subset of some region of G_{n+1} and of some domain of V_n and of D, and (2) the sum of all the domains of V_n is a connected domain D_n containing M. By Theorems 79 and 80 of Chapter I, the set of all points common to all the sets of the sequence $D_1, D_2, D_3, \cdots$ is a compact continuum, and it contains M and lies in D.

A modification of this argument proves this theorem for a space which satisfies Axioms 0 and 1 and is locally arcwise connected.

THE UNIVERSITY OF TEXAS

[1] Presented to the Society, February 24, 1940.

[2] American Mathematical Society Colloquium Publications, vol. 13, New York, 1932. All references are to this book.

ON THE MEAN VALUES OF AN ANALYTIC FUNCTION[1]

RAPHAEL M. ROBINSON

This note contains improvements on the results in two recent papers by Nehari.[2]

The first paper shows that if $f(z)$ is regular for $|z|<1$, and if the mean of $|f(z)|$ on the circle $|z|=r$ is less than or equal to 1 for each $r<1$, then the mean of $|f(z)|^2$ on $|z|=r$ is less than or equal to 1 for $r\leqq 6^{-1/2}$. We shall show that the conclusion is true for $r\leqq 2^{-1/2}$, but not always for a larger value of r. *More generally, we shall show that the mean of* $|f(z)|^p$ *on* $|z|=r$ *is less than or equal to* 1 *for* $r\leqq p^{-1/2}$ (*where* $p>1$ *is an integer*), *and that this result is the best possible.*

It will be sufficient to prove that if $g(z)$ is a function which is regular for $|z|\leqq 1$ and different from 0 for $|z|<1$, and such that the mean of $|g(z)|$ on $|z|=1$ is less than or equal to 1, then the mean of $|g(z)|^p$ on $|z|=r$ is less than or equal to 1 for $r\leqq p^{-1/2}$. For suppose $0<R<1$, and put

$$g(z)=f(Rz)\colon\prod_{n=1}^{n}\frac{z-\alpha_n}{1-\bar{\alpha}_n z},$$

where $\alpha_1, \alpha_2, \cdots, \alpha_n$ are the zeros of $f(Rz)$ in $|z|<1$. We note that $|g(z)|=|f(Rz)|$ for $|z|=1$, while $|g(z)|>|f(Rz)|$ for $|z|<1$. The function $g(z)$ evidently satisfies the above conditions. From the conclusion that the mean of $|g(z)|^p$ on $|z|=r$ is less than or equal to 1 for $r\leqq p^{-1/2}$, we see that the mean of $|f(Rz)|^p$ on $|z|=r$ is not greater than 1 for $r\leqq p^{-1/2}$, or that the mean of $|f(z)|^p$ on $|z|=r$ is not greater than 1 for $r\leqq Rp^{-1/2}$. The desired result follows by letting $R\to 1$.

We have to show that from the hypothesis $(1/2\pi)\int_0^{2\pi}|g(e^{i\theta})|\,d\theta\leqq 1$ the conclusion

$$\frac{1}{2\pi}\int_0^{2\pi}|g(re^{i\theta})|^p d\theta\leqq 1, \qquad \text{for } r\leqq p^{-1/2},$$

follows. Now since $g(z)\neq 0$ for $|z|<1$, we may put $g(z)=h(z)^2$, where $h(z)$ is regular for $|z|<1$. If we put

$$h(z)=\sum_{n=0}^{\infty}a_n z^n, \qquad h(z)^p=\sum_{n=0}^{\infty}c_n z^n,$$

[1] Presented to the Society December 2, 1939.

[2] Comptes Rendus de l'Académie des Sciences, Paris, vol. 206 (1938), pp. 1943–1945; vol. 208 (1939), pp. 1785–1787. My results were obtained during a summer (1939) spent at Stanford University. The two papers mentioned were called to my attention by Professor Szegö.

and use the well known formula

$$\frac{1}{2\pi}\int_0^{2\pi} |h(re^{i\theta})|^2 d\theta = \sum_{n=0}^{\infty} |a_n|^2 r^{2n},$$

we see that the hypothesis becomes $\sum_{n=0}^{\infty} |a_n|^2 \leqq 1$, while in a similar manner the desired conclusion becomes

$$\sum_{n=0}^{\infty} |c_n|^2 r^{2n} \leqq 1, \qquad \text{for } r \leqq p^{-1/2}.$$

Now $c_n = \sum a_{n_1} a_{n_2} \cdots a_{n_p}$, where the sum extends over all sets $(n_1, n_2, \cdots, n_p)$ of integers not less than 0 whose sum is n. Hence by the Cauchy-Schwarz inequality, we have

$$|c_n|^2 \leqq \sum 1 \cdot \sum |a_{n_1}^2 a_{n_2}^2 \cdots a_{n_p}^2|,$$

where the sums have the same range as before. Now $\sum 1$ is the number of ways of distributing n units among p terms, and hence is not greater than p^n, which is the number of ways of distributing n *different* things among p sets. Hence for $r \leqq p^{-1/2}$ we have

$$|c_n|^2 r^{2n} \leqq \sum |a_{n_1}^2 a_{n_2}^2 \cdots a_{n_p}^2|,$$

and therefore

$$\sum_{n=0}^{\infty} |c_n|^2 r^{2n} \leqq \left(\sum_{n=0}^{\infty} |a_n|^2 \right)^p \leqq 1.$$

The theorem is not true for $r > p^{-1/2}$. For if $\epsilon > 0$ and we put $f(z) = (1+\epsilon z)^2/(1+\epsilon^2)$, then the hypothesis of the theorem is satisfied. On the other hand, $f(z)^p = (1+p\epsilon z + \cdots)^2/(1+p\epsilon^2 + \cdots)$, so that the mean of $|f(z)|^p$ on $|z| = r$ is $(1+p^2\epsilon^2 r^2 + \cdots)/(1+p\epsilon^2 + \cdots)$, which is greater than 1 if $pr^2 > 1$ and ϵ is sufficiently small. This negative conclusion is true also for non-integral values of p; but we have been able to prove the positive statement only for integral values of p.

We turn now to the second paper mentioned. In this, it is proved that if $f(0) = 0$ and if the mean of $|f(z)|$ along each radius of the unit circle is not greater than 1, then the mean of $|f(z)|$ along $|z| = r$ is less than or equal to 1 for $r \leqq \frac{1}{2}$, but not always for a larger value of r. The negative part of the statement is immediate, the counter-example being $f(z) = 2z$. *We shall show that the hypothesis* $f(0) = 0$ *is unnecessary, and that the stronger statement that the mean of* $|f(z)|^2$ *along* $|z| = r$ *is less than or equal to* 1 *for* $r \leqq \frac{1}{2}$ *is also true.*

We prove first the following result. *If the mean of* $|F(z)|^2$ *on* $|z| = r$

is not greater than 1 *for* $r<1$, *then the mean of* $|F'(z)|^2$ *on* $|z|=r$ *is less than or equal to* 1 *for* $r\leqq\frac{1}{2}$. If we put $F(z)=\sum_{n=0}^{\infty}b_nz^n$, we have only to prove that

$$\sum_{n=1}^{\infty}n^2|b_n|^2r^{2(n-1)}\leqq\sum_{n=0}^{\infty}|b_n|^2 \qquad \text{for } r\leqq\tfrac{1}{2}.$$

Since $n\leqq 2^{n-1}$ for all positive integral values of n, we see that $nr^{n-1}\leqq 1$, so that the inequality is true. (The result is not correct for $r>\frac{1}{2}$; counterexample, $F(z)=z^2$.)

We suppose now that the mean of $|f(z)|$ along each radius of the unit circle is not greater than 1, and put

$$F(z)=\int_0^z f(\zeta)d\zeta.$$

Since the integral may be taken along a radius, we see that

$$|F(z)|\leqq 1, \qquad \text{for } |z|<1.$$

Hence the mean of $|F(z)|^2$ on $|z|=r$ is certainly not greater than 1 for any $r<1$. Therefore the mean of $|F'(z)|^2=|f(z)|^2$ on $|z|=r$ is not greater than 1 for $r\leqq\frac{1}{2}$.

UNIVERSITY OF CALIFORNIA

REMARKS ON A NOTE OF MR. R. WILSON AND ON RELATED SUBJECTS[1]

G. SZEGÖ

Introduction. *Let $w(x)$ be a nonnegative weight function on the interval $-1 \leqq x \leqq +1$, and let the integral*

$$\int_{-1}^{+1} \log w(x) \cdot (1 - x^2)^{-1/2} dx = \int_0^{\pi} \log w(\cos\theta) d\theta \tag{1}$$

exist in the sense of Lebesgue.

If $\{p_n(x) = k_n x^n + \cdots\}$ denotes the orthonormal set of polynomials associated with $w(x)$, we have

$$\lim_{n\to\infty} \max_{-1\leqq x\leqq +1} |p_n(x)|^{1/n} = 1, \tag{2}$$

and[2]

$$\lim_{n\to\infty} k_n^{1/n} = 2. \tag{3}$$

In 1921 I found[3] the following asymptotic formula for the orthogonal polynomials $p_n(x)$ for $n \to \infty$, holding for x not on the segment $[-1, +1]$:

$$\lim_{n\to\infty} z^n p_n(x) = \Delta(z) \tag{4}$$

where $2x = z + z^{-1}$, $|z| < 1$, and $\Delta(z)$ is a certain analytic function regular and nonzero for $|z| < 1$. Of course, $\Delta(z)$ depends on the weight function $w(x)$. The formula (4) holds uniformly for

$$|z| \leqq r, \quad r < 1.$$

From this result the formulas (3) and, by an additional elementary remark (cf. below (9)), (2) follow immediately. Also it furnishes (cf. OP, p. 302, Theorem 12.7.1):

$$\lim_{n\to\infty} 2^{-n} k_n = \pi^{-1/2} \exp\left\{-\frac{1}{2\pi}\int_0^{\pi} \log w(\cos\theta) d\theta\right\}. \tag{5}$$

---

[1] Presented to the Society, February 24, 1940.

[2] Concerning the notation see my book *Orthogonal Polynomials* (American Mathematical Society Colloquium Publications, vol. 23, 1939). Hereafter this book will be referred to as OP.

[3] G. Szegö, *Über die Entwickelung einer analytischen Funktion nach den Polynomen eines Orthogonalsystems*, Mathematische Annalen, vol. 82 (1921), pp. 188–212; p. 191. Cf. also OP, p. 290, Theorem 12.1.2.

Mr. R. Wilson devoted a recent note[4] to the proof of the relation (2), or rather of the following relation:

$$\lim_{n\to\infty} \max_{-1\leqq x\leqq 1} \left| k_n^{-1} p_n(x) \right|^{1/n} = \tfrac{1}{2}, \tag{6}$$

which is, on account of (3), equivalent to (2). His argument is based on certain results of Mr. Shohat,[5] which are incidentally consequences of the asymptotic formula (4). The conditions used by Shohat are more restrictive than the existence of (1).

By applying a classical theorem of Poincaré on recurrence formulas, Shohat proves that

$$\lim_{n\to\infty} \left| p_n(x) \right|^{1/n} = \left| z \right|^{-1}, \tag{7}$$

where x and z have the same meaning as before and $|z|<1$. (From this, (3) follows for $x=\infty$ or $z=0$.) Based on (7), a proof of (2) or (6) can easily be arranged. Wilson (loc. cit., p. 191) prefers, however, to use another theorem on recurrence formulas due to Perron.[6]

In the present note, I give first a very simple direct approach to (2), (3), and (6), assuming the existence of (1). Naturally the deeper result (5) requires more refined methods.

Further, we deal with the following related result of Shohat (loc. cit., pp. 34–36): *Let $w(x)\geqq 0$ be an arbitrary weight function on the interval $-1\leqq x\leqq +1$. Employing the former notation, the relations*

$$2^{-n}k_n = O(1), \qquad \lim_{n\to\infty} 2^{-n}k_n \text{ exists} \tag{8}$$

are equivalent. Of course, this means that the second relation follows from the first one. The proof, given below, is essentially Shohat's argument; we found it, however, convenient and possible to eliminate every reference to the theory of continued fractions used by Shohat.

Finally, by means of the deeper result (5) we show that the existence of the integrals (1) is not only *sufficient* but also *necessary* for the relations (8). More precisely: *Let $w(x)$ be a nonnegative weight*

[4] R. Wilson, *A note on the asymptotic properties of orthogonal polynomials*, this Bulletin, vol. 45 (1939), pp. 190–192.

[5] See J. Chokhatte (Shohat), *Sur le développement de l'intégrale $\int_a^b [p(y)/(x-y)]dy$ en fraction continue et sur les polynômes de Tchebycheff*, Rendiconti del Circolo Matematico di Palermo, vol. 47 (1923), pp. 25–46; cf. in particular pp. 43–44.

[6] The following objection can be made to his argument. In the present case the coefficients of the recurrence formula contain x as a parameter. However, $\max |p_n(x)|$, $-1\leqq x\leqq +1$, will be in general attained for an $x=x(n)$ which varies with n. Therefore, a generalization of Perron's theorem is needed here stating the *uniform* existence of the limit involved.

function on the interval $-1 \leqq x \leqq +1$, *and let* $p_n(x) = k_n x^n + \cdots$ *have the same meaning as before. If* $2^{-n} k_n = O(1)$, *the integrals* (1) *exist.*

Proof of (2), (3), (6). It is well known that

$$\max_{-1 \leqq x \leqq +1} \left| k_n^{-1} p_n(x) \right| \geqq 2^{1-n}, \tag{9}$$

so that for (6) it suffices to show that in $-1 \leqq x \leqq +1$

$$k_n^{-1} \left| p_n(x) \right| < A \cdot 2^{-n} (1 + \epsilon)^n; \tag{10}$$

here $\epsilon > 0$ is arbitrary and A depends only on $w(x)$ and ϵ.

Let $z = re^{i\phi}$, $0 \leqq r < 1$. By use of the inequality between the arithmetic and geometric mean of a function,[7] we find that

$$\begin{aligned} \pi^{-1} &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} w(\cos\theta) \left| \sin\theta \right| \left\{ p_n(\cos\theta) \right\}^2 d\theta \\ &\geqq \frac{1-r}{1+r} \frac{1}{2\pi} \int_{-\pi}^{+\pi} w(\cos\theta) \left| \sin\theta \right| \left\{ p_n(\cos\theta) \right\}^2 \\ &\qquad \cdot \frac{1-r^2}{1 - 2r\cos(\theta - \phi) + r^2} d\theta \\ &\geqq \frac{1-r}{1+r} \cdot \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{+\pi} \log\left[w(\cos\theta) \left| \sin\theta \right| \right] \right. \\ &\qquad \left. \cdot \frac{1-r^2}{1 - 2r\cos(\theta - \phi) + r^2} d\theta \right\} \\ &\qquad \cdot \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{+\pi} \log\left\{ p_n(\cos\theta) \right\}^2 \right. \\ &\qquad \left. \cdot \frac{1-r^2}{1 - 2r\cos(\theta - \phi) + r^2} d\theta \right\}. \end{aligned} \tag{11}$$

The harmonic function

$$\log \left| z^n p_n(\tfrac{1}{2}(z + z^{-1})) \right|^2 = 2\Re \log \left\{ z^n p_n(\tfrac{1}{2}(z + z^{-1})) \right\} \tag{12}$$

is regular for $|z| < 1$, and it has logarithmic singularities at $z = e^{\pm i\alpha_\nu}$, if $\cos\alpha_\nu$, $\nu = 1, 2, \cdots, n$, denote the roots of $p_n(x)$. Therefore, the second exponential expression in (11) becomes

$$\exp\left\{ \log \left| z^n p_n(\tfrac{1}{2}(z + z^{-1})) \right|^2 \right\} = \left| z \right|^{2n} \left| p_n(\tfrac{1}{2}(z + z^{-1})) \right|^2.$$

Consequently,

[7] See, for instance, G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. 1, Berlin, 1925, pp. 54, 210, Problem 79.

$$|z|^n |p_n(\tfrac{1}{2}(z+z^{-1}))| < \pi^{-1/2}\left(\frac{1+r}{1-r}\right)^{1/2} \cdot \exp\left\{\frac{1}{2\pi}\int_0^\pi |\log[w(\cos\theta)\sin\theta]|\,d\theta \cdot \frac{1+r}{1-r}\right\}. \tag{13}$$

For $z=0$ this results in the inequality

$$2^{-n}k_n < \pi^{-1/2}\exp\left\{\frac{1}{2\pi}\int_0^\pi |\log[w(\cos\theta)\sin\theta]|\,d\theta\right\}.$$

On the other hand. for $x=\cos\theta$,

$$k_n^{-2} = \min \int_{-1}^{+1} w(x)(x^n + \cdots)^2 dx \leqq \int_{-1}^{+1} w(x)(2^{1-n}\cos n\theta)^2 dx < 2^{2-2n}\int_{-1}^{+1} w(x)dx,$$

so that $2^{-n}k_n$ remains between two positive bounds. From here (3) follows. Choosing $|z| = 1-\delta$ in (13) where $\delta>0$ is sufficiently small, and applying the maximum-principle in the corresponding ellipse of the x-plane, we obtain (10) which implies (6) and also (2).

Proof of the equivalence of the relations (8) (Shohat's theorem). Let (see OP, p. 41, Theorem 3.2.1)

$$p_n(x) = k_n\bar{p}_n(x) = k_n(x^n - S_n x^{n-1} + \cdots), \tag{14}$$

$$\bar{p}_n(x) = (x - c_n)\bar{p}_{n-1}(x) - \left(\frac{k_{n-2}}{k_{n-1}}\right)^2 \bar{p}_{n-2}(x), \tag{15}$$

$$b_{2n} = -\frac{k_{n-1}}{k_n}\frac{p_n(-1)}{p_{n-1}(-1)} > 0, \quad b_{2n+1} = -\frac{k_{n-1}}{k_n}\frac{p_{n-1}(-1)}{p_n(-1)} > 0, \tag{16}$$

$$n = 1, 2, 3, \cdots; \ \bar{p}_{-1}(x) = 0.$$

Obviously,

$$k_0^{-2}k_n^2 = (b_2 b_3 \cdots b_{2n}b_{2n+1})^{-1}, \tag{17}$$

$$c_n + 1 = b_{2n} + b_{2n-1}, \ S_n = \sum_{\nu=1}^{2n} b_\nu - n, \ n = 1, 2, 3, \cdots; b_1 = 0. \tag{18}$$

Since

$$c_n\int_{-1}^{+1}\{p_{n-1}(x)\}^2 w(x)dx = \int_{-1}^{+1} x\{p_{n-1}(x)\}^2 w(x)dx, \tag{19}$$

we have $|c_n|<1$, so that from (18) $0<b_n<2$ follows (except $b_1=0$).

If $T_n(x)$ denotes the polynomial of Tchebichef, we write

$$T_n(x) = h_0p_0(x) + h_1p_1(x) + \cdots + h_np_n(x); \tag{20}$$

then

$$h_\nu = \int_{-1}^{+1} T_n(x)p_\nu(x)w(x)dx,$$

$$|h_\nu| \leqq \int_{-1}^{+1} |p_\nu(x)|\, w(x)dx \leqq \left\{\int_{-1}^{+1} w(x)dx\right\}^{1/2}, \tag{21}$$

$$\nu = 0, 1, 2, \cdots, n.$$

Comparing the coefficients of x^n and x^{n-1} in (20), we find $2^{n-1}=h_nk_n$, $0=-h_nk_nS_n+h_{n-1}k_{n-1}$, so that

$$S_n = 2^{1-n}h_{n-1}k_{n-1}, \qquad |S_n| \leqq 2^{1-n}k_{n-1}\left\{\int_{-1}^{+1} w(x)dx\right\}^{1/2}.$$

Now, we define $U_n=2b_n-1$, $-1<U_n<3$ (except $U_1=-1$). Then, as $n\to\infty$,

$$\sum_{\nu=1}^{2n} U_\nu = 2S_n = 2^{1-n}k_{n-1}\cdot O(1),$$

$$\prod_{\nu=2}^{2n+1}(1+U_\nu)^{-1} = \prod_{\nu=2}^{2n+1}(2b_\nu)^{-1} = 2^{-2n}k_0^{-2}k_n^2 = (2^{-n}k_n)^2\cdot O(1). \tag{22}$$

Assuming $2^{-n}k_n=O(1)$, both expressions (22) remain bounded.

Let c be a positive constant such that

$$u^{-2}\{u - \log(1+u)\} > c, \qquad -1 < u < 3. \tag{23}$$

Thus $cU_n^2<U_n-\log(1+U_n)$, so that $\sum_{\nu=1}^{2n}U_\nu^2=O(1)$, that is, $\sum U_n^2$ is convergent. The same holds for

$$\begin{aligned}\sum_{n=1}^{\infty} U_{2n}U_{2n+1} &= \sum_{n=1}^{\infty}\{4b_{2n}b_{2n+1} - 2(b_{2n}+b_{2n+1}) + 1\}\\ &= \sum_{n=1}^{\infty}\{4b_{2n}b_{2n+1} - 2(b_{2n}+b_{2n-1}) + 1\}\\ &\quad + \sum_{n=1}^{\infty} 2(b_{2n-1} - b_{2n+1})\\ &= \sum_{n=1}^{\infty}\left\{4\left(\frac{k_{n-1}}{k_n}\right)^2 - 2(S_n - S_{n-1}) - 1\right\}\\ &\quad + \sum_{n=1}^{\infty} 2(b_{2n-1} - b_{2n+1}).\end{aligned} \tag{24}$$

The last series is convergent since $U_n \to 0$ or $b_n \to \frac{1}{2}$. Therefore,

$$\sum_{n=1}^{\infty} \left\{ 4\left(\frac{k_{n-1}}{k_n}\right)^2 - 2(S_n - S_{n-1}) - 1 \right\} \tag{25}$$

is convergent. Applying this result to the polynomials $(-1)^n p_n(-x) = k_n(x^n + S_n x^{n-1} + \cdots)$ associated with the weight function $w(-x)$ on $[-1, +1]$, we obtain the convergence of

$$\sum_{n=1}^{\infty} \left\{ 4\left(\frac{k_{n-1}}{k_n}\right)^2 - 2(-S_n + S_{n-1}) - 1 \right\}, \tag{26}$$

or that of $\sum_{n=1}^{\infty}(S_n - S_{n-1})$. This is equivalent to the existence of $\lim_{n\to\infty} S_n$, or to the convergence of $\sum U_n$. This, together with the convergence of $\sum U_n^2$, implies the convergence of the product $\prod(1+U_n)^{-1}$ or the existence of $\lim_{n\to\infty} 2^{-n}k_n$.

Proof of the equivalence of the conditions (8) to the existence of the integrals (1). The relations (8) are equivalent to the fact that for every polynomial $q(x) = x^n + \cdots$ of the nth degree with the highest term x^n

$$\int_{-1}^{+1} \{q(x)\}^2 w(x)dx > c \cdot 2^{-2n} \tag{27}$$

holds, where $c > 0$ is independent of n. Let $\epsilon > 0$; we have

$$\int_{-1}^{+1} \{q(x)\}^2 (w(x) + \epsilon)dx > c \cdot 2^{-2n}. \tag{28}$$

The minimum of the left-hand side is $\{k_n(\epsilon)\}^{-2}$, where $k_n(\epsilon)$ denotes the highest coefficient of the orthonormal polynomial of the nth degree associated with $w(x)+\epsilon$. For this weight function, the integral condition (1) is satisfied so that according to (5)

$$\lim_{n\to\infty} 2^{-n}k_n(\epsilon) = \pi^{-1/2} \exp\left\{ -\frac{1}{2\pi}\int_0^{\pi} \log\,[w(\cos\theta) + \epsilon]d\theta \right\}. \tag{29}$$

Therefore,

$$\pi \exp\left\{ \pi^{-1}\int_0^{\pi} \log\,[w(\cos\theta) + \epsilon]d\theta \right\} \geqq c,$$

$$\int_0^{\pi} \log\,[w(\cos\theta) + \epsilon]d\theta \geqq c'$$

where c' is independent of ϵ. Now, $\log(\alpha+\beta) < \log\beta + 1$ for $0<\alpha<\frac{1}{2}<\beta$; so if $0<\epsilon<\frac{1}{2}$,

$$\int_{w(x)>1/2} \log[w(\cos\theta)+\epsilon]d\theta < \int_{w(x)>1/2} \{\log w(\cos\theta)+1\}d\theta,$$

hence

$$\int_{w(x)\leqq 1/2} \log[w(\cos\theta)+\epsilon]d\theta > c''$$

where c'' is independent of ϵ. The same inequality holds if the integration is extended over the set $0<\eta\leqq w(x)\leqq\frac{1}{2}$. But for a decreasing sequence of bounded (negative) functions the operations of integration and passing to the limit as $\epsilon\to+0$ are interchangeable; consequently,

$$\int_{\eta\leqq w(x)\leqq 1/2} \log w(\cos\theta)d\theta \geqq c'',$$

and since this is true for all $\eta>0$, the integral (1) exists.

STANFORD UNIVERSITY